

# Resource-Constrained Low-Complexity Video Coding for Wireless Transmission

**Ukhanova, Ann**

*Publication date:*
2012

*Document Version*
Publisher's PDF, also known as Version of record

# Resource-Constrained Low-Complexity Video Coding for Wireless Transmission

Anna Ukhanova

31 August 2012

Coding and Visual Communication Group
Department of Photonics Engineering
Technical University of Denmark
Building 343
2800 Kgs. Lyngby
DENMARK

DTU Fotonik
Institut for Fotonik

To my parents

# Abstract

Constrained resources like memory, power, bandwidth and delay requirements in many mobile systems pose limitations for video applications. Standard approaches for video compression and transmission do not always satisfy system requirements. In this thesis we have shown that it is possible to modify and optimize conventional algorithms in order to convert them into low-complexity solutions and satisfy system constraints.

We have studied low-complexity approaches for video compression without motion estimation. We have proposed scalable (progressive) solutions for video compression with low memory consumption based on image coding standards. Scalability aspects were studied for distributed video coding as well. We have compared temporal scalability for distributed and scalable video coding and provided recommendations for the choice of one of these solutions based on the system requirements. Another comparison regarded power consumption for distributed video coding and H.264/AVC standard. We also proposed a scalable-to-lossless extension of transform domain Wyner-Ziv codec that allows bit savings compared to lossless coding by standard algorithms. Scalability aspects were also studied in perspective of video quality. We proposed a new metric for objective quality assessment that considers frame rate.

As many applications deal with wireless video transmission, we performed an analysis of compression and transmission systems with a focus on power-distortion trade-off. We proposed an approach for rate-distortion-complexity optimization of upcoming video compression standard HEVC. We also provided a new method allowing decrease of power consumption on mobile devices in 3G networks. Finally, we proposed low-delay and low-power approaches for video transmission over wireless personal area networks, including 60GHz fiber-wireless link.

# Resumé

Begrænsede ressourcer som hukommelse, energi, båndbredde og forsinkelse medfører begrænsninger for video-applikationer i mobile systemer. Standardmetoder til videokompression og transmission opfylder ikke altid systemkravene. I denne afhandling har vi vist, at det er muligt at modificere og optimere konventionelle algoritmer og dermed konvertere dem til lavkompleksitetsløsninger, der tilfredsstiller systemets begrænsninger.

Vi har undersøgt tilgange med lav kompleksitet for videokompression uden bevægelsesestimation. Vi har foreslået skalerbare løsninger med lavthukommelsesforbrug baseret på billed-kodningsstandarder. Skalerbarhedsaspekter blev også undersøgt for distribueret videokodning. Vi har sammenlignet tidsmæssig skalerbarhed med distribueret og skalerbar videokodning og givet anbefalinger til valget af en af disse løsninger baseret på systemkravene. I en anden sammenligning betragtes energiforbrug for distribueret videokodning og H.264/AVC standard. Vi har også foreslået en skalerbar-til-tabsfri udvidelse af transformationsdomæne Wyner-Ziv codec, der tillader bitbesparelse i forhold til tabsfri kodning med standardalgoritmer. Skalerbarhedsaspekter blev også undersøgt med henblik på videokvalitet. Vi har foreslået en ny metrik for objektiv kvalitetsvurdering, der inkluderer frame rate.

Da der er mange videotransmissionsapplikationer, har vi analyseret videokommunikationssystemer med fokus på et trade-off mellem energi og forvrængning. Vi har foreslået en strategi for rate-forvrængning-kompleksitets optimering af en kommende videokompressionsstandard HEVC. Vi har præsenteret en ny metode, der gør det muligt at reducere energiforbruget på mobile enheder i 3G-net. Endelig har vi foreslået energibesparende metoder med lave forsinkelser til videotransmission via trådløse personlige netværk, herunder et 60GHz fiber-trådløst link.

# Acknowledgements

This thesis presents a selection of the research carried out during my Ph.D. studies at the Coding and Visual Communication Group, DTU Fotonik, Technical University of Denmark, in the period from September 2009 to August 2012.

It is my pleasure to thank those people who made this thesis possible. First, I wish to thank my advisor, Prof. Søren Forchhammer for all his thoughtful guidance, insightful criticism and immense knowledge. His oral and written comments were always perceptive and appropriate. He has been a demanding advisor throughout my Ph.D. studies, but he has always given me a great deal of freedom to pursue my own research directions. I would like to thank the reviewers of my thesis - Prof. Patrick Le Callet, Dr. Eskil Faber and Prof. Knud Larsen - for their valuable comments and suggestions.

I would like to acknowledge the co-authors of my papers, and particularly Eugeniy Belyaev for his guidance and contribution in joint works and for showing me a great example of being a scientist.

Also, I would like to thank my colleagues from the Coding group and express my gratefulness to Jari Korhonen for his patience, perseverance in obtaining results and numerous discussions of our joint work. I would like to thank in particular Claire Mantel and Matteo Salmistraro for proof-reading of the thesis and their helpful advices.

I wish to thank Otto Mønsteds Fond and Oticon Fonden for supporting my participation in the conferences and my research stay abroad during my Ph.D. studies.

I am grateful to Simone Milani for offering me an opportunity for an external stay in the University of Padua, Italy, where I spent 4 months during my Ph.D. studies. I learned a lot from him and the work we had

been performing together has opened to me new insights on my thesis.

I would like to express my gratitude to my friends in Italy that made my stay there an unforgettable experience. Most of all, I wish to thank Luisa Ciatto, Bruno and Marta Zamarin, and Francesco Artuso for their care, love and continuous support during my months in Italy. They made me feel there as in my own family and I am extremely grateful for that! I am also very thankful to Paolo Crivellari and Lucia Lazzaretto for their friendship, hospitality and care.

I would like to express my appreciation to my friends, Nino Burini and Federica Genovese who were sources of enthusiasm, laughter and fun during many delightful evenings spent together.

I wish to thank my brothers, my entire family and my friends all over the world who were in contact with me during these years for not forgetting me and providing their faith in me.

I cordially thank Marco Zamarin for his infinite support and care. He was giving me inspiration and additional forces to work on the thesis. I would like to express lots of gratitude for him for proof-reading, sensible amendments and stylistic suggestions that helped to improve the presentation of my research work.

Finally, I would like to thank my parents for their love and endless support. They gave me forces to complete this work and made it easier for me to make crucial decisions, among them to start and finish my Ph.D. studies far away from them.

Anna Ukhanova

# Ph.D. Publications

This Ph.D. project has resulted in 11 peer-reviewed publications in total, including 2 journal publications and 9 conference contributions. The publications are listed below.


**[1]** **A. Ukhanova**, J. Korhonen, S. Forchhammer "Objective assessment of the impact of frame rate on video quality", *International Conference on Image Processing (ICIP)*, Lake Buena Vista (FL), USA, 2012.

**[2]** **A. Ukhanova**, E. Belyaev, L. Wang, S. Forchhammer "Power consumption analysis of constant bit rate video transmission over 3G networks", *Computer Communications Journal*, Vol. 35, Iss. 14 (Special issue: Wireless Green Communications and Networking), 2012.

**[3]** A. Lebedev, T. Pham, M. Beltrán, X. Yu, **A. Ukhanova**, R. Llorente, I. Monroy, S. Forchhammer, "Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band", *Optics Express Journal*, Vol. 19, Iss. 26, 2011.

**[4]** A. Lebedev, T. Pham, M. Beltrán, X. Yu, **A. Ukhanova**, L. Deng, N. Gonzalez, R. Llorente, I. Monroy, S. Forchhammer, "Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band", *European Conference on Optical Communication (ECOC)*, Geneva, Switzerland, 2011.

**[5]** L. Wang, **A. Ukhanova**, E. Belyaev, "Power consumption analysis of constant bit rate data transmission over 3G mobile wireless networks", *11th International Conference on Telecommunications for Intelligent Transport Systems (ITST)*, St. Petersburg, Russia, 2011.

**[6]** **A.Ukhanova**, A. Sergeev, S. Forchhammer, "Extending JPEG-LS for low-complexity scalable video coding", *IS&T/SPIE Electronic Imaging*, San Francisco (CA), USA, 2011.

**[7]** **A.Ukhanova**, S. Forchhammer, "Low-complexity JPEG-based progressive video codec for wireless video transmission", *International Conference on Ultra Modern Telecommunications (ICUMT)*, Moscow, Russia, 2010.

**[8]** X. Huang, **A. Ukhanova**, A. Veselov, S. Forchhammer, M. Gilmutdinov, "Scalable-to-Lossless Transform Domain Distributed Video Coding", *Multimedia Signal Processing (MMSP)*, St.-Malo, France, 2010.

**[9]** **A. Ukhanova**, E. Belyaev, S. Forchhammer, "Encoder power consumption comparison of Distributed Video Codec and H.264/AVC in low-complexity mode", *The 18th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Bol, Croatia, 2010.

**[10]** E. Belyaev, A.Turlikov, **A. Ukhanova**, "Low-latency video transmission over high-speed WPANs based on low-power compression", *IEEE Wireless Communications & Networking Conference (WCNC)*, Sydney, Australia, 2010.

**[11]** X. Huang, **A. Ukhanova**, E. Belyaev, S. Forchhammer, "Temporal scalability comparison of the H.264/SVC and Distributed Video Codec", *International Conference on Ultra Modern Telecommunications (ICUMT)*, St. Petersburg, Russia, 2009.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation

The $21^{st}$ century is marked by a rapid development in many spheres of Information and Communications Technologies (ICT). Within just few years there has been a significant growth in amount of internet users that has almost doubled from 2006 to 2011, and in amount of mobile-cellular subscriptions worldwide that have almost reached 6 billion. At the same time internet bandwidth grew exponentially from 11000 GBit/s in 2006 to almost 80000 GBit/s in 2011, and mobile broadband subscriptions have reached almost 1.2 billion [12]. Together with internet communications and mobile systems there is also a vast development of related applications. Among these video applications are taking a significant place. According to Cisco, by the end of 2011 mobile video traffic was 52 % of all data traffic. Moreover, they predict that by 2016 over 70 % of the world's mobile data traffic will be video [13]. Thus, video applications are rising in popularity and becoming ubiquitous. A good examples of popular video applications are videoconference services such as Skype [14] or video sharing websites like YouTube [15]. At the same time, the amount of mobile subscriptions is also growing very fast. Therefore bandwidth demand due to data, especially video, is increasing. It means that in these conditions video compression is essential as uncompressed data transmission requires a much higher bandwidth, which is often not feasible for current transmission technologies.

### 1.1.1 State-of-the-art video coding standards

In the last decades there have been published many international standards, such as Advanced Video Coding (H.264/AVC) [16, 17], Scalable Video Coding (H.264/SVC) [18,19], Multiview Video Coding (MVC) [20, 21] and some other video compression algorithms like Distributed Video Coding (DVC) [22]. As they were developed in different times and for different needs, these algorithms have different features and differ in their performance and algorithmic properties. For example, DVC represents a low-complexity solution for the encoder while Scalable Video Coding (SVC) provides scalable bit streams. If applied in mobile systems, these algorithms have to be adapted to handheld devices that have constrained resources such as memory and power. Additional limitations include delay requirements, especially in real-time video communications, and constrained bandwidth. There is no ideal solution that satisfies all these requirements, but there is an ongoing research in the direction of video codecs adjustment and development, and a number of improvements have been suggested over time.

The H.264/AVC standard is the state-of-the-art in video coding and is widely used, e.g. for video streaming, High-Definition Television (HDTV) broadcasts and in Blue-ray discs. Its successor, the High Efficiency Video Coding (HEVC) [23], is being standardized nowadays. It promises to be a more efficient solution in terms of compression performance and it is supposed to be adapted to higher resolution videos.

### 1.1.2 Challenges in video codec design

Creation of new and more efficient, but at the same time more complex, standards does not mean that more simple solutions cannot be useful as well. Existing video communication systems and relative video applications are so different in nature and required features that it seems unfeasible to develop a general solution that suits all needs. For example, high-complexity solutions do not fit applications that need low power consumption. The state-of-the-art scalable video codec H.264/SVC is quite complex which can restrict its use in low-power systems such as video surveillance or video streaming where the scalability feature can be very convenient. Therefore, it might be necessary to have an alternative solution that provides comparable performance but with simpler imple-

mentation in terms of complexity and power consumption costs [6, 7].

It is also important to have effective ways to compare the performance of different video compression schemes. The most complicated case is the one of scalable schemes that produce compressed sequences with various frame rates, spatial resolutions or fidelity. While there exist conventional methods of comparing quality (fidelity), like Peak Signal-to-Noise Ratio (PSNR), it is still an open question how to compare video sequences with different temporal and spatial resolutions. Novel objective metrics try to find an effective solution to this problem [24–27].

### 1.1.3 Wireless video applications

Video transmission takes a significant part of the overall data transmission in mobile systems. Wireless video transmission schemes are based on various mobile networks such as 3G or high-speed Wireless Personal Area Networks (WPANs) [28–30]. These schemes can be naturally divided into compression and transmission parts in which optimization and adaptation are applied either jointly or separately. The overall power consumption depends on the complexity of both parts, which are influencing each other. Overall power consumption of video transmission systems can be adjusted by power management systems. Cross-layer optimization allows controlling transmission parameters in order to achieve desired power consumption level. At the encoder side it is represented by a rate-control algorithm that tries to minimize the overall distortion under given bit rate constraints ideally taking both encoder power consumption and latency requirements into account [10].

### 1.1.4 Goals of the thesis

The main focus of this Ph.D. project is on efficient low-complexity solutions for video communication systems under constrained resources. This work considers low-complexity and scalable compression schemes, quality assessment for scalable video coding and applications of video compression in communication systems.

The goals of this thesis are to show that it is possible to modify, adjust and optimize conventional approaches in video compression and transmission in order to convert them into low-complexity solutions operating under specific constraints. Among additionally studied features, we paid

a close attention to scalability issues due to the growing heterogeneity of the wireless networks and capabilities of client terminals.

## 1.2 Structure of the thesis

The Ph.D. studies resulted in 11 peer-reviewed journal and conference contributions [1–11]. 9 of them [1–3,6–11] are included in this thesis (see Appendix A).

The publications are within the fields of distributed video coding, scalable video coding, low-complexity approaches for video communication systems, power consumption analysis and visual quality assessment.

The remainder of this thesis is organized as follows: in Section 1.3 we provide an overview of Ph.D. publications that represent the base of this thesis. These publications are divided in two groups: "Video compression systems" (Chapter 2) and "Video transmission systems" (Chapter 3).

Chapter 2 gives an introduction to video compression systems with a focus on low-complexity and scalable solutions. It begins with an overview of general scheme of video codec. Further we discuss characteristics and properties of compression algorithms and describe comparison criteria. Image and video codecs are presented afterwards. In this chapter we also discuss applications with constrained resources that require the use of specific compression algorithms, and describe quality assessment for scalable video data.

In Chapter 3 aspects of wireless video transmission are discussed. First, we describe modern communication systems and application scenarios for systems with constrained resources such as delay and power consumption. We take a closer look at 3G networks and WPANs as examples for application of video compression to wireless networks. We consider that complexity relates to the system power consumption either in terms of battery capacity or hardware processing power. Therefore, we propose cross-layer optimization for low-delay low-complexity video transmission in 3G networks and WPANs, including joint fiber-wireless transmission in the 60 GHz band. Then we discuss rate-distortion-complexity trade-off and propose a method for complexity control for the upcoming HEVC standard.

Finally, conclusions and outlooks for the work presented in this thesis are given in Chapter 4.

## 1.3 Description of Ph.D. publications

### 1) Video compression systems

- **Encoder power consumption comparison of Distributed Video Codec and H.264/AVC in low-complexity mode [9]**

  In this paper we have compared a DVC encoder [31] with an alternative low-complexity encoder - baseline profile of H.264/AVC working in differential frame coding mode with Context-Adaptive Variable-Length Coding (CAVLC) as entropy encoder. The comparison criterion was based on the power that the encoder consumes in order to provide a given PSNR value. We designed a simple model for relative power consumption of H.264/AVC and DVC and provided an equation that shows the dependency of the power consumption gain on implementation efficiency of CAVLC and Low-Density Parity-Check (LDPC) coder and coding efficiency of the compared algorithms. Rate-distortion performance combined with estimated power consumption was used for evaluation of PSNR vs. power consumption relationship. Results showed that, for the most relevant visual quality range of 30 - 40 dB, DVC with LDPC coder allows to decrease the power consumption of entropy coding about 15 - 60% compared to CAVLC in H.264/AVC with differential frame coding. However, only the kernel of the video codecs has been compared. Moreover, DVC utilizes H.264/AVC encoder for intra/key frames. This leads to such an increase in the complexity that it may be preferable to use H.264/AVC in differential coding mode in systems that require low encoding complexity.

- **Temporal scalability comparison of the H.264/SVC and Distributed Video Codec [11]**

  In this work we have analyzed the efficiency of temporal scalability in DVC [22] and H.264/SVC [18]. For comparison we have used DISCOVER DVC [31], our implementation of DVC codec, and the H.264/SVC reference software [32]. We presented scalability implementations in DVC and SVC solutions, discussed the differences and presented the test conditions. We provided also a comparison of complexity and required memory size, and of rate-distortion performance for the discussed approaches. Results showed that with the availability of the temporal scalability, the appropriate encoding method (scalable or distributed video coding) should be chosen based on the memory constraints.

- **Scalable-to-Lossless Transform Domain Distributed Video Coding [8]**

  In this work we presented a scalable-to-lossless distributed video codec as an extension of lossy Transform Domain Wyner-Ziv (TDWZ) distributed video codec [33] with feedback based on Discrete Cosine Transform (DCT). Scalable-to-lossless solutions may be used for high-quality applications where lossless is desirable, but the system cannot (at least, efficiently) guarantee the resources for lossless coding. We proposed to apply reversible integer DCT to provide lossless coding. We also described a new backward adaptive coding scheme and illustrated potential for improvement of scalable-to-lossless TDWZ as well as feasibility of efficient scalable-to-lossless coding using reversible integer DCT. Experimental results showed that the proposed scalable-to-lossless TDWZ video codec can outperform alternatives based on the JPEG2000 standard. As for lossless coding efficiency, the proposed scalable-to-lossless TDWZ codec can save between 5 - 13% in terms of bit stream size compared to lossless coding by JPEG-LS, JPEG2000 and H.264 Intra frame coding. Compared with low-complexity Inter frame lossless coding schemes (i.e. JPEG-LS Diff and JPEG2000 Diff), the proposed scalable-to-lossless TDWZ video codec gave better or comparable performance for sequences with complex motion, but not competitive results for the mostly static sequences.

- **Extending JPEG-LS for low-complexity scalable video coding [6]**

  In this paper we proposed a simple low-complexity solution for two-step scalable lossless compression based on JPEG-LS. We compared our solution with the leading image and video coding standards supporting scalability - JPEG2000 and H.264/SVC - in low-complexity modes, suitable for wireless video applications. We presented the constraints motivated by wireless applications and adapted configuration parameters of the codecs to these conditions. We provided the rate-distortion performance evaluation of the proposed and competitor algorithms for the task of real-time wireless high-quality low-complexity video transmission. The comparison showed that the proposed scheme can provide better performance than standard solutions for specific video contents at high rates.

- **Low-complexity JPEG-based progressive video codec for wireless video transmission [7]**

  In this paper we took a look at the enhancement of video coding in the case of memory and complexity constraints. First, we showed the possibility of tuning the JPEG2000 algorithm for different image types like natural or synthetic images when the tiling option is used in order to provide low memory consumption. Then, we proposed our simplified version of progressive JPEG algorithm adapted for video compression under specific constraints, for which existing image and video compression algorithms, suitable for general conditions, fail to give the best performance. We slightly modified the standard JPEG algorithm in almost all coding blocks ranging from domain forming procedure to Run Lengths encoding procedure and Huffman coding. Compared to the full-featured JPEG2000 coding system with adjusted parameters, we got a simple solution with reduced complexity and memory consumption that allows providing good compression performance on High-Definition (HD) computer graphics, while for other contents the compression/quality ratio decreases slightly.

- **Objective assessment of the impact of frame rate on video quality [1]**

  In this paper, we studied the impact of frame rate on video quality and proposed a novel objective quality metric that takes temporal resolution into account. Using the results from a previous subjective study [34], we estimated the Perceptual PSNR (PPSNR) and found a function that allows predicting these values by using PSNR, frame rate and a content dependent parameter that can be easily obtained from spatial and temporal characteristics of the video. In order to evaluate the performance of the proposed metric, our metric has been compared with two other objective metrics, VQMTQ [26] and STVQM [27]. We observed that, on average, the proposed metric is capable of predicting the subjective preferences more accurately than those two metrics.

### 2) Video transmission systems

- **Low-latency video transmission over high-speed WPANs based on low-power compression [10]**

  In this paper we discussed latency-constrained video transmission over high-speed WPANs. We analyzed end-to-end distortion and showed that for its minimization it is sufficient to minimize the quantization distortion only. We further presented a near-optimal video source rate control algorithm based on the MINMAX quality criterion. The proposed rate control is constructed in one-pass mode, therefore it does not contribute much to the overall encoder power consumption. Practical results for video transmission using H.264/AVC standard over 60 GHz channel showed that in "good" channel conditions the low-power version of H.264/AVC encoder with rate-control provides lossless video bit rate equal to 1.5 Gbps, which allows saving channel throughput. In "bad" channel states, the proposed rate control algorithm provides adaptation to varying channel conditions and guarantees acceptable video quality for the given channel throughput.

- **Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band [3]**

  (This paper is an extended version of the conference paper [4])

  In this paper we presented the study of high quality HD compressed video transmission over a 60 GHz fiber-wireless network. Our experiments and simulations demonstrated that there is a trade-off between the distortion introduced by the source (due to lossy compression) and distortion introduced by the channel. We achieved high delivery quality for a given link budget by joint optimization of physical layer parameters of a Radio over Fiber (RoF) link (power levels, distance) and codec parameters (quantization, error-resilience tools) for H.264/AVC in a low-complexity low-delay configuration. Employment of error-resilience tools of H.264/AVC with simplified settings showed a greater robustness against impairments that occur in 60 GHz fiber-wireless channel.

- **Power consumption analysis of constant bit rate video transmission over 3G networks [2]**

  (This paper is an extended version of the conference paper [5])

  This paper presents an analysis of the power consumption of video data transmission with constant bit rate over 3G mobile wireless networks. We proposed a model for power consumption in the most consuming state of the Radio Resource Control (RRC) and showed how power consumption is related to the packet size and transmission interval. The comparison with the reference model based on data rate showed that the proposed model provides a better approximation of the experimental results. We discussed how the power consumption depends on the transmission parameters and explained how it can be reduced by optimizing the power management policy. Namely, we proposed a method of parameter selection for the 3GPP transition state machine that allowed to decrease power consumption on the mobile device in case of constant bit rate data transmission. Further, we investigated few particular cases with limitations for the amount of signaling traffic, buffer size and buffering latency. We extended our analysis to the case of constant bit rate video transmission and discussed the gain in

power consumption vs. PSNR when applying our proposed method instead of the one currently used in 3G. We also presented the possibility of performing power consumption management based on the requirements for the video quality.

# Chapter 2

# Video compression systems

## 2.1 Introduction

Video data is widely used in many spheres of our life. Digital broadcasting brings high-quality TV signal to our houses, CD, DVD and Blu-ray discs allow to store movies, Internet services such as YouTube offer millions of videos, videoconference services like Skype let us see a conversational partner in real time and modern medical devices are also often equipped with videocameras.

Video compression systems allow to reduce the amount of bits needed for data storage or transmission. In contrast to compressed video signals, uncompressed video data require larger storage space or transmission bandwidth. An uncompressed video sequence of Common Intermediate Format (CIF) resolution (352×288) and 30 frame per second (fps) in 4:2:0 color format has a bit rate of approximately 37 Mbps while for full-color 4:4:4 uncompressed High-Definition Television (HDTV) video sequence with 60 fps we need approximately 3 Gbps. A movie in this format with a length of 90 minutes would require over 2000 Gigabytes (GB) of data space.

At the same time, 3G networks provide on average bit rates from few hundreds kbps to few Mbps. Typical DVD bit rates vary from 4 to 10 Mbps and digital video broadcasting achieves bit rates up to 19 Mbps. The maximum available DVD disc capacity is slightly higher than 17 GB.

As can be seen from the numbers above, existing channel throughputs

and digital media device capacities are not high enough for transmission and storage of uncompressed video data. Most of the modern computers do not allow real-time HDTV uncompressed video playing either. An essential solution to this problem is video data compression, which allows using channel and storage resources more efficiently.

Many compression standards and algorithms, published in the recent time, allow efficient solutions for video encoding. These solutions differ in the provided features, performance and characteristics. Each particular video compression task requires an individual approach in the choice of appropriate solution. For example, video applications designed for low-power mobile devices require low-complexity methods for video compression due to the limited battery capacity and hardware capabilities. Another example of applications for wireless video transmission demonstrates the need for scalable solutions due to the heterogeneity of the clients in communication systems. Thus, requirements for specific features of compression algorithms make the task of video codec design more complex and challenging.

### 2.1.1 General scheme

A general scheme of video compression system is shown in Fig. 2.1. It consists of an encoder and a decoder without communication channel. The scheme of a joint video compression and transmission system is described in Section 3.1. The general approach to video coding in compression standards includes the following steps: reduction of the color correlation, exploitation of the redundancies (spatial and temporal) that are related to the similarities and predictability of data, quantization and entropy coding.

#### Color transform

Red, Green and Blue (RGB) components are correlated among themselves. In order to reduce this correlation and allow higher compression, color space conversion can be applied to RGB components of the video sequence. This conversion is often called RGB to YUV color space transform, however, it is important to note that this name can refer to different models, namely YUV, YCbCr models and some others. Below we provide the transform from RGB to YCbCr, which is used in JPEG and

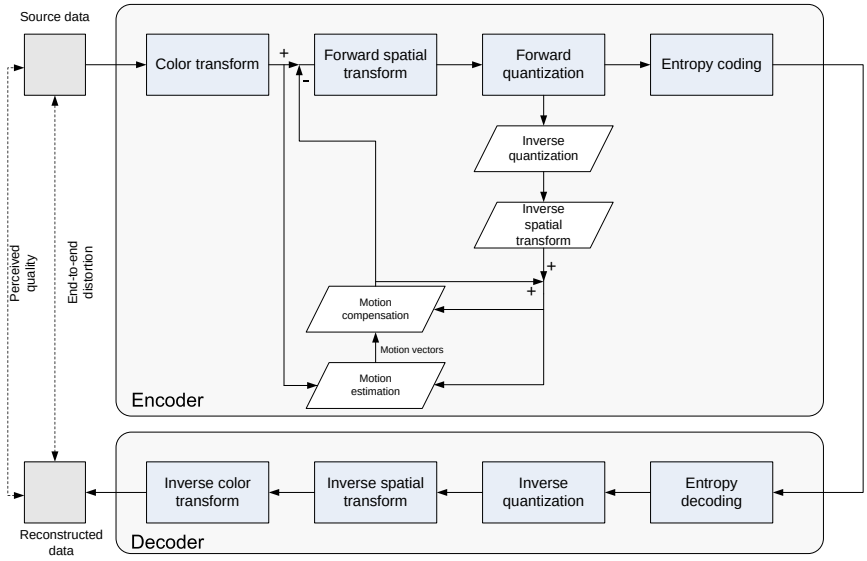



**Figure 2.1:** Video compression scheme

MPEG coding standards. Here Y denotes luminance component and Cb and Cr are the two chrominance components [35].

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix} \quad (2.1)$$

### Spatial transforms

Various transforms like Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) are used in video coding standards for reducing spatial redundancy, caused by the correlation of nearby pixels. These transforms allow concentrating most of the signal energy in a few transformed coefficients, typically corresponding to low-frequency components. The forward and inverse DCT of a 2-D $8 \times 8$ block are defined

by the following formulas [35]:

$$S_{uv} = \frac{1}{4} C_u C_v \sum_{i=0}^{7} \sum_{j=0}^{7} s_{ij} \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \tag{2.2}$$

$$s_{ij} = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C_u C_v S_{uv} \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}, \tag{2.3}$$

where

$$C_u C_v = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u, v = 0 \\ 1 & \text{otherwise,} \end{cases} \tag{2.4}$$

$S_{uv}$ is the transformed $(u, v)$ DCT coefficient,
$s_{ij}$ is the value of the pixel at the position $(i, j)$ in the block.

The DWT scheme is described below in the example of JPEG2000 standard. DWT can be alternatively performed using the lifting scheme [35]. Two wavelet filters are used for DWT in JPEG2000 (see Fig. 2.2): a high frequency filter $h_0(n)$ and a low frequency filter $h_1(n)$. First, the wavelet transform is applied on rows, followed by downsampling with a factor of two; then on columns with downsampling with a factor of two afterwards as well. The four obtained matrices $HH_0$, $HL_0$, $LH_0$, $LL_0$ correspond to the filtering by $h_0(n)$ over rows and columns, $h_0(n)$ over rows and $h_1(n)$ over columns, $h_1(n)$ over rows and $h_0(n)$ over columns and $h_1(n)$ over rows and columns. Then DWT can be performed over low-pass matrix $LL_0$ once again, resulting in matrices $HH_1$, $HL_1$, $LH_1$, $LL_1$. If DWT is applied to the original signal $k$ times, the result will be consisting of $3k + 1$ diminishing matrices.

### Motion estimation

Adjacent frames can also be very similar and changes in them are typically caused by the motion of objects or of the camera, except for scene cuts. It is therefore possible to reduce temporal redundancy between frames by using motion compensated prediction, consisting of Motion Estimation (ME) and Motion Compensation (MC). The most widely used block-based ME first chooses the best matching block in the reference frame $F_{ref}$ and finds the displacement between the current block

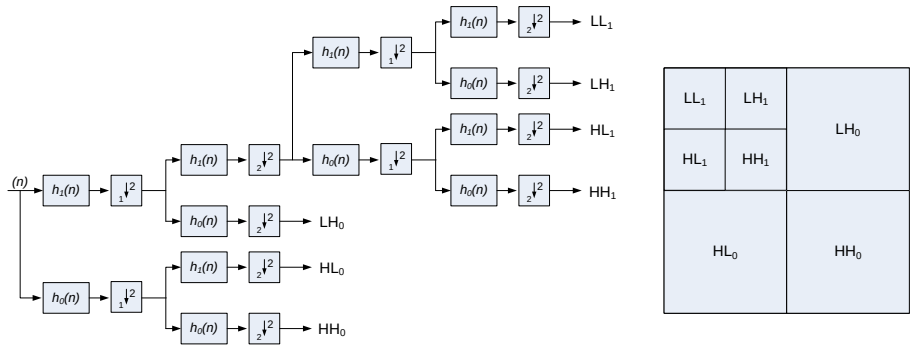



**Figure 2.2:** Discrete wavelet transform [36]

$F_{cur}(x, y)$ and the best matching block $F_{ref}(x+m_x, y+m_y)$ within the defined search range. Afterwards, the residual (difference) block $\Delta F(x, y)$ between the blocks of current frame $F_{cur}$ and reference frame $F_{ref}$ is calculated as follows:

$$\Delta F(x, y) = F_{cur}(x, y) - F_{ref}(x + m_x, y + m_y). \tag{2.5}$$

Further, the residual $\Delta F(x, y)$ is encoded together with the so-called motion vectors $m_x$ and $m_y$. If differential frame coding mode is used, then $m_x$ and $m_y$ are set to 0.

Being an efficient approach for video coding, ME at the same time gives a significant contribution to the complexity of the encoder that grows with the increase of the search range.

### Quantization

Quantization is the next step in the video compression process. It allows representing a large range of coefficient values with a smaller set at the price of reduced fidelity. Quantization matrices for scalar quantization are constructed taking into account the fact that the human visual system is more sensitive to low frequencies, therefore high frequencies values are usually quantized more. The quantization block in video compression scheme exploits visual (subjective) redundancy or, in other words, visual irrelevancy. The loss caused by quantization is irreversible and afterwards lossless reconstruction is no longer possible. Two parts of

the decoder - inverse quantization and inverse spatial transform - are included in the encoder scheme (see Fig. 2.1) so that an identical data is used in the encoder and decoder for the next frame and drifting is prevented.

### Entropy coding

Finally, an entropy coder is applied to the quantized coefficients. Among entropy coding schemes used in image and video compression standards are Variable-Length Coding (VLC) (e.g. Huffman coder that produces codes with optimal length [37] or Context-Adaptive Variable-Length Coding (CAVLC) [17]) and adaptive arithmetic coder that considers varying symbol statistics (e.g. Context-Adaptive Binary Arithmetic Coding (CABAC) [38]).

In the conventional compression scheme the encoder is more complex than the decoder.

### Decoder scheme

The scheme of the decoder can be represented as an inverse scheme of the encoding process. Decoded quantized values are scaled by inverse quantization. Then inverse spatial transform allows to reconstruct the residuals that are afterwards combined with the result of the prediction. The decoded picture is used for the prediction of following frames and can be displayed after inverse color space transform. Typically, it is the decoder that is standardized. This provides flexibility to the encoding scheme and allows different implementations.

### General scheme applications

The scheme described above is common for many video compression algorithms like ITU-T Recommendation H.261 [39], where it was first introduced, and its successors MPEG-1 [40], H.262 MPEG-2 [41], H.263 [42], MPEG-4 Part 2 (Visual) [43], H.264/MPEG-4 Part 10 (AVC) [16] and High Efficiency Video Coding (HEVC) [23]. Given that slight variations and modifications are possible, the flexibility of this scheme allows to adjust its blocks to the given conditions and by this achieve relevant gains.

### 2.1.2 Comparison criteria

Even if most algorithms follow the video compression scheme presented in the previous section, they differ in their characteristics and performance. Compression algorithms can be divided into lossless, providing reconstruction of original data without errors, and lossy, when some compression artifacts are introduced. Other parameters distinguishing algorithms are their performance, encoder and decoder complexity, end-to-end delay, scalability, rate-control options, robustness to data losses and errors, Region Of Interest (ROI) coding, possibility of random access etc. Below we present several main comparison criteria for video codecs: compression ratio, quality and complexity.

#### Compression ratio

The compression ratio is one of the most obvious and important criterion. It shows the ratio between compressed and uncompressed data sizes, i.e. compression efficiency of the algorithm. It can be computed by the following formula:

$$\text{Compression Ratio} = \frac{\text{Compressed data size}}{\text{Uncompressed data size}}. \tag{2.6}$$

For the evaluation of the compression performance of the algorithm, bit rate is often used instead of compression ratio. Bit rate for $N$ compressed frames of the considered video sequence can be computed as follows:

$$\text{Bit Rate} = \frac{\text{Frame rate}}{\text{N}} \sum_{i=0}^{N-1} \text{Compressed data size}(i). \tag{2.7}$$

#### Quality

Another important criterion for comparison of video compression schemes is the quality of the reconstructed video data or distortion. The methods for quality measurement are divided into two categories: subjective and objective measurements. Objective quality assessment analyzes the characteristics of the video signal while subjective quality assessment is

based on visual evaluation by human observers. Subjective quality assessment provides the most reliable results that are called ground truth, as quality is a subjectively perceived feature by nature. Subjective quality is often represented by Mean Opinion Score (MOS). The most widely used approach to collect MOS is to ask test subjects to rate the quality of test stimulus using five point scale ranging from one (bad) to five (excellent) [44]. As test subjects may interpret subjective quality as well as rating scales differently, it is often necessary to remove outliers from the raw test results. However, subjective tests can be influenced not only by human factors, but also by inadequate test arrangements and limited test material. Moreover, subjective tests require more resources and are not suitable for some applications, e.g. real-time quality monitoring. Therefore, there is an ongoing development of objective quality metrics that correlate as good as possible with the human perception of the video sequences.

One of the most widely used metrics, the Peak Signal-to-Noise Ratio (PSNR), is computed between two $m \times n$ 8 bpp monochrome images $X$ and $Y$ by the following formula:

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i,j) - Y(i,j)]^2}. \tag{2.8}$$

However, PSNR does not correlate perfectly with the subjective opinion [45], neither is it directly applicable for comparison of video sequences with different spatial or temporal resolution, so more accurate objective quality metrics are needed. Quality metrics have been steadily evolving in the recent time [46], and several metrics taking frame rate impact into account have been proposed, such as QM [25], VQMTQ [26] and STVQM [27]. However, they have a number of disadvantages, such as poor performance on low frame rates or complex computation of metric parameters.

We have proposed a novel objective metric for quality evaluation of the sequences with different frame rates [1]. Our metric is based on PSNR, frame rate and a content dependent parameter that can easily be obtained from slightly modified spatial and temporal activity indices of the video sequence [47]. On average, the proposed metric shows a slightly better performance than other metrics [1]. A modification of this

metric and its performance on different databases is discussed further in Section 2.5.

### Complexity

The third important criterion for the evaluation of compression algorithm is codec complexity that plays an important role for the devices with constrained resources. It is typically divided into encoder and decoder complexity, and can be estimated in terms of amount of elementary operations, memory or power consumption. It is possible to control the complexity of the communication system using complexity management strategies [48]. Such strategies for mobile devices allow to find a desired trade-off between compression performance and computational complexity. The complexity of the compression algorithm is subject to limitations due to the device hardware constraints and/or limited battery capacity. On the one hand, low-complexity solutions allow real-time processing; on the other hand, they help to decrease the overall power consumption of the device.

### Rate-distortion performance

As the goal of video compression is to compress the video signal with maximum quality provided the bit rate constraint, two of the above-mentioned criteria - bit rate and distortion - are typically combined into rate-distortion performance. It indicates the resulting distortion $D$ (e.g. PSNR) of the video signal with a given rate $R$. It is a widely spread method for comparison of codec performance and it is used in benchmark tests [49]. Distortion can be estimated using one of the quality metrics. Video quality metrics that take temporal or spatial scalability dimensions into account allow comparing video sequences with different frame rates and resolutions. In the case of bit rate constraints, scalable codecs allow reducing bit rate by higher quantization or by reduction in frame rate or resolution. An optimal solution in this case can be found by objective quality metrics applicable for multidimensional scalability.

## 2.2 Algorithm characteristics

### 2.2.1 Scalability

Video compression systems vary not only in rate-distortion performance, but in main features and characteristics. One of the important features for wireless video transmission is scalability. It allows to encode the video data only once and decode it at different quality levels. Three main types of scalability are distinguished: spatial, temporal and quality or Signal-to-Noise Ratio (SNR). Spatial scalability allows video decoding at multiple spatial resolutions, temporal scalability deals with the reduction of frame rate, while quality scalability keeps the same spatio-temporal resolution but enables decoding with lower fidelity. All these types can be also merged together in the so-called combined scalability (Fig. 2.3) [50]. The encoder forms the bit stream divided in several layers - base layer and enhancement layers. The base layer contains the information about the video stream at the lowest resolutions for all three dimensions - quality, temporal and spatial. Decoding of each additional enhancement layer allows improving the overall video quality of the reconstructed sequence at the expense of an increased bit rate.

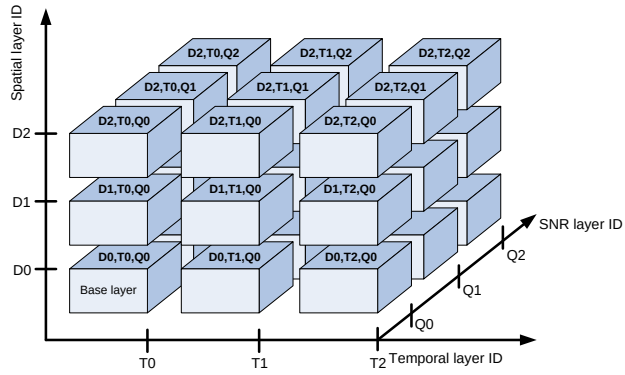



**Figure 2.3:** Combined scalability [51]

### 2.2.2 Power consumption

For many mobile systems and handheld devices power consumption is an important issue. The power consumption of a compression algorithm is directly connected to its complexity and memory consumption. Memory and computational constraints can be caused by the commercial use of these video compression systems. Algorithm complexity becomes another parameter for comparison of various systems. When speaking about complexity of video compression systems, it is important to note whether it refers to the encoder or decoder complexity. The complexity of one encoder can also vary depending on particular configurations. For example, the state-of-the-art codec H.264/AVC [16, 17] has several encoding profiles with different computational complexities. Two of the most complex and consuming blocks in the system are the ME and MC blocks [52, 53]. These blocks are not used in the Intra low-complexity profile. In order to achieve higher performance using low-complexity configurations, it is possible to use differential frame coding, i.e. ME with zero search radius [54]. In this way comparable performance and encoder complexity between H.264/AVC and Distributed Video Coding (DVC) can be achieved [9].

### 2.2.3 Memory consumption

Memory constraints due to power consumption restrictions or due to commercial reasons also play a role in the choice of an optimal compression scheme. The amount of memory needed at the encoder and decoder sides can be reduced by using tiling [7], i.e. division of the frames into non-overlapping rectangular blocks that are compressed independently. As only one tile is usually kept in the memory, it prevents an efficient use of MC and ME or other methods exploiting temporal and spatial redundancies. This usually leads to a performance decrease, however a solution based on detection of static regions between two consecutive frames, which we proposed in [10], is still possible.

## 2.3 Image and video codecs

The idea of combining spatial image compression and temporal motion compensation laid the basis for efficient video compression schemes. In

the last few decades there has been a fast evolution in image and video compression systems from simple image codecs like JPEG [55,56] that is, however, still widely used nowadays, till upcoming standard HEVC [23] (its release is scheduled for 2013), that seems to be a promising solution for high-resolution video data. Below we provide brief codec descriptions and give an overview of this evolution.

### 2.3.1 Image coding

#### JPEG standard

The popular image compression standard JPEG dates back to 1990s. It was issued few years after the Joint Photographic Expert Group (JPEG), responsible for developing image coding standards, was organized. JPEG algorithm refers to the group of algorithms based on DCT. After color transform from RGB to YUV color space, DCT is applied to the non-overlapping $8 \times 8$ pixel blocks, one color component at a time. The top-left corner value of each transformed block is a DC coefficient, while other 63 coeffients are referred to as AC. Transformed coefficients are quantized using quantization matrices, defined in JPEG standard. Quantized DC coefficient is encoded as a difference between the current and previous DC coefficients. Quantized AC values are rearranged in a so-called Zigzag order to collect most of zero values together. Run Length Encoding (RLE) is used to efficiently encode groups of zeros which are afterwards encoded with Huffman coding. Huffman tables are applied separately to AC and DC coefficients of luma and chroma components. All Huffman tables are defined in JPEG standard.

JPEG is a universal algorithm that has good rate-distortion performance on average. Despite being developed in the early 90s, it is still widely used nowadays in the web and in digital cameras. It can be used in many applications, and if an application has specific requirements or works with specific contents, then JPEG can be modified in order to achieve better performance in the particular case. For example, computer graphics image statistics differ from natural image ones. This image property can be used for adaptation of the JPEG algorithm.

If JPEG is applied to each frame of a video sequence, it can be used for video compression. This approach is informally referred to as Motion JPEG (MJPEG) and it is used e.g. in some digital and IP cameras.

Recent work shows that MJPEG can also be applied for scalable secure video communications [57]. We have shown [7] that if a progressive scheme instead of a standard JPEG is used for MJPEG, it provides to the bit stream a scalability property, which is important for wireless video transmission. Our progressive encoder has a lower complexity than progressive mode of JPEG, which does the encoding in multiple scans.

### JPEG-LS standard

Even though JPEG was able to operate in a lossless mode based on Differential Pulse Code Modulation (DPCM), this mode was never widely adopted. In 1999 the JPEG standard was complemented with a lossless compression scheme called JPEG-LS. It has a similar name, but it is based on a completely different approach - predictive technique. Based on the context, one of the two possible modes is chosen - "Regular" or "Run". In "Run" mode pixels having equal (or almost equal in near-lossless encoding) values are encoded together using run length coding. In "Regular" mode instead the algorithm chooses one of the possible probability distributions and uses it to encode the prediction error with a Golomb code. JPEG-LS is much faster than another algorithm from the JPEG family - JPEG2000 [58, 59] - that also allows lossless compression. In addition to lossless compression JPEG2000 also provides scalability. However, we have shown that it is possible to modify JPEG-LS to implement a low-complexity scalable solution [6].

### JPEG2000 standard

The compression standard JPEG2000 appeared in the JPEG family in 2000. This time the standard is based on DWT and it aimed to surpass JPEG in performance and flexibility. As mentioned above, big advantages of JPEG2000 against JPEG are scalability and the possibility of lossless compression. In addition it has ROI coding and higher error resilience thanks to data encoding in independent blocks. The general compression scheme is similar to that of JPEG. After color space transform (choice available between reversible and irreversible), the full image or its tiles (if tiling option is on) are transformed by means of DWT. Transformed coefficients are further scalar-quantized and the result is sent to an entropy coding block called Embedded Block Coding with Optimal

Truncation (EBCOT) [60]. Its extension Motion JPEG2000 [61] allows applying it to video sequences. JPEG2000 is computationally more complex than JPEG and JPEG-LS, and it has a number of configuration parameters that need to be tuned in order to achieve the best performance. In [7] we have studied the application of JPEG2000 scheme for low-memory encoding. In order to achieve it the tiling option needs to be used, and other configuration parameters - such as number of DWT levels and codeblock size - need to be adjusted.

#### JPEG XR

In 2009 new still image compression standard JPEG XR was published [62, 63]. It utilizes a scheme similar to JPEG, but supports higher bit depths, uses smaller blocks for transforms and has some other changes in entropy coding and other compression blocks. It provides both lossy and lossless coding and is designed for continuous tone photographic images. However, up to now it did not receive as wide acceptance as JPEG.

### 2.3.2 Video coding

#### MPEG and H.26x standards

Modern video coding standards such as MPEG and H.26x have two modes - intra-frame coding where spatial correlation within a frame is used, and inter-frame coding in which temporal correlation between frames is exploited. If a frame is encoded using intra-prediction only, it is referred to as I frame. If a reference frame from one direction is used for prediction, the encoded frame is referred to as P frame. Finally, in the B frame bidirectional prediction is used. The approach described above makes encoders quite complex as they use ME and MC, transform domain coding and entropy coding, while keeping the decoder side relatively light. Such a scheme fits well the needs of video broadcasting when video data has to be encoded only once, but decoded at various client terminals many times. However, such solution is not suitable, for example, for wireless video sensor networks, where the system needs light-weight encoder instead.

**H.264/AVC standard.** The state-of-the-art H.264/AVC standard follows the general scheme shown on Fig. 2.1. Additionally, a deblocking filter can be applied after inverse transform in encoding procedure in order to remove the blocking effect and achieve better ME results. H.264/AVC standard has several complexity profiles, specifying the set of used coding tools. For low-complexity low-delay applications the Baseline profile was developed. It supports I and P frame types, intra prediction and inter prediction from a single reference, the basic $4 \times 4$ integer transform and CAVLC as an entropy coder [64]. It has low computational complexity but provides lower performance than the High or Main profiles that allow using more sophisticated tools.

**Scalable extension of H.264/AVC standard.** H.264/AVC has also a scalable extension H.264/SVC [18, 19], which provides scalable bit streams in temporal, spatial, quality or combined dimensions. As it is based on H.264/AVC standard, most of the main operations are the same. H.264/SVC can produce a base layer which is backward compatible with H.264/AVC. H.264/SVC temporal scalability is also based on H.264/AVC concept of hierarchical prediction structure using P and B frames (Fig. 2.4). Regarding temporal scalability, SVC differs only in temporal layer signaling.

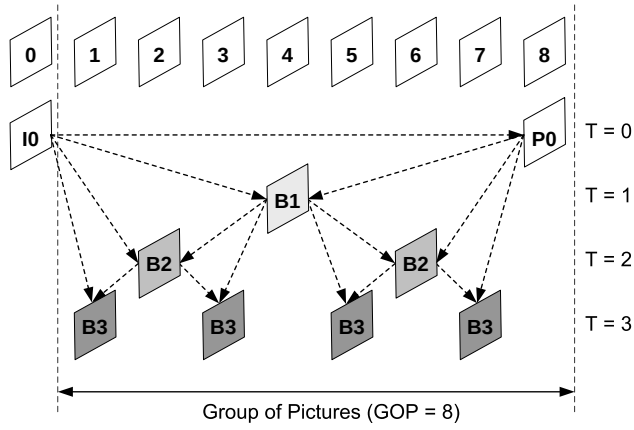



**Figure 2.4:** Temporal scalability provided by P and B frames [51]

**Upcoming HEVC standard.** H.264/AVC successor - HEVC - is following a similar compression scheme. However, HEVC introduces a new data structure - Coding Unit (CU) tree. The CU is the basic coding unit in HEVC, so it appears to be the analog of a macroblock in H.264/AVC. While macroblock partition sizes in H.264/AVC vary from $4\times4$ to $16\times16$, CU sizes can go up to $64 \times 64$. Such wide range allows efficient encoding of large homogeneous areas that are more typical in high resolution sequences. The partition depth in the configuration of HEVC defines the smallest block size that can be used for encoding. In HEVC Test Model (HM) [65] partition depth can vary from 1 to 4. For example, if CU is defined as $64 \times 64$ block, and maximum partition depth is set to 3, CU splitting can go down to $16 \times 16$ blocks. We say "can" as it is not mandatory that the whole frame will be encoded with blocks $16 \times 16$. It means that on partition depth levels 3 and 2 (i.e. $16 \times 16$, $32 \times 32$ block sizes in this case) the sum of the rate-distortion costs for four CUs of that size will be compared with the rate-distortion cost of the encoding CU at the following depth level (e.g. four CUs of size $16 \times 16$ at depth level 3 will be compared to single CU of size $32 \times 32$ at depth level 2, and four CUs of size $32 \times 32$ at depth level 2 will be compared to single CU of size $64 \times 64$ at depth level 1) and the decision for partition will be taken based on the smallest rate-distortion cost. Fig. 2.5 presents an example of hierarchical CU structure described above. This method, even though

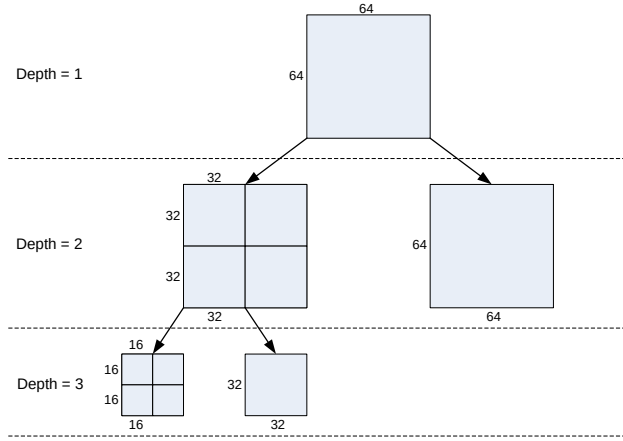



**Figure 2.5:** CU tree structure

providing efficient compression, has a high computational complexity as the encoder has to try encoding CUs on all levels in all possible configurations. Thus, in order to use HEVC in power-constrained devices, complexity management has to be performed at the encoder side.

#### Distributed video coding

In contrast to the conventional video coding approaches described above where source statistical dependencies are exploited at the encoder side, a new paradigm for video compression - distributed coding - was recently developed. It is based on the results obtained by Slepian and Wolf (for lossless coding) [66] and Wyner and Ziv (for lossy coding) [67]. DVC encodes frames independently using intra-frame coding, but decodes them conditionally using "side information" from previously decoded frames. Therefore, the complexity of the compression system is shifted to decoder. DVC represents a novel architecture for low-power video compression where an efficient coding may be achieved by light-weight video encoder. DVC looks like a promising solution for mobile systems and wireless transmission, even though its decoder complexity is rather high. It has high robustness to packet losses due to architecture, and transcoding allows to make the decoding for mobile devices simple too.

An architecture implemented in DISCOVER DVC [31] allows temporal scalability of the bit stream. We have compared it with temporal scalability in H.264/SVC [11] and concluded that the preference towards one or another solution depends on the memory consumption requirements. We have also extended transform domain DVC to a scalable-to-lossless version [8]. This solution may be desirable for high-quality applications, in which the system cannot always guarantee resources for lossless compression.

### 2.3.3 Codecs evolution

As it can be seen from the overview of international standards and coding approaches, video compression has experienced a rapid development in the past two decades and this technological evolution continues nowadays. Video compression standards are used in various services and applications, employed in wireless networks and mobile terminals. This development required taking many aspects into account. As noticed by

Netravali and Limb [68] "(...) the application of picture coding to transmission channels is an economic trade-off in system design, balancing picture quality, circuit complexity, bit rate and error performance." In other words, the idea of practical video coding design is to achieve an optimal trade-off between rate and resulting quality, provided the limitations for maximum complexity and delay [69]. These limitations often pose restrictions for the deployment of many compression methods.

Originally, the simplest video compression method was an independent compression of each picture. However, the average performance of such an approach is relatively low compared to full-featured video codecs. In fact, an exploitation of temporal redundancy can significantly improve the performance. A combination of both prediction and transformation laid the basis for MPEG and H.26x standards that kept the block-based structure. As within the time the complexity constraints have been eased, it was possible to add more coding modes in order to improve the compression efficiency. For example, compared to MPEG-4 and H.263 standards, that support only the blocks of size $16 \times 16$ and $8 \times 8$, H.264 allows to process partitions with block sizes $16 \times 16$, $16 \times 8$, $8 \times 16$, $8 \times 8$, $8 \times 4$, $4 \times 8$ and $4 \times 4$.

The HEVC standardization committee is going further in increasing the number of coding modes and proposes to have more than 30 different modes for intra prediction and block sizes go up to $64 \times 64$. Due to these and other modifications, HEVC provides an average bit rate savings compared to H.264/AVC for about 39% for random access and 44% for low delay configurations [70]. As current video technologies allow to capture High-Definition (HD) video content and continue to evolve towards Ultra High-Definition (UHD) resolutions, there is a demand for new generation video compression technologies that can efficiently process these contents. The upcoming standard HEVC promises to have certain advances in video communications in terms of supported resolutions, compression performance, efficiency-complexity trade-offs, robustness to errors and flexibility for usage in various applications and services.

Scalable functionality of video codecs was required in heterogeneous wireless environments. Simple ways of achieving scalability are simulcasting and transcoding. However, these methods are more resource consuming than the creation of a scalable bit stream during the encod-

ing process. Temporal scalability was supported to some extent already in MPEG-1 with the introduction of I, P and B frame coding types. An example of temporal scalability provided by P and B frames is shown in Fig. 2.4.

In MPEG-2 spatial scalability was also introduced. However, scalability in MPEG-2 and MPEG-4 was causing a significant loss in compression efficiency. H.264/SVC allowed to solve this problem. The bit rate increase was up to 10% only compared to a non-scalable solution. It was also supporting a backward compatible H.264/AVC base layer. In addition to adaptive delivery in lossy transmission environments, SVC layered structure allows a combination with Unequal Error Protection (UEP) providing in this way error resiliency.

The codec scheme mentioned above requires much more complex encoder than decoder. A novel DVC paradigm allowed to shift the encoder complexity to the decoder. Additional advantages of DVC are robustness to channel errors due to joint source-channel coding (e.g. the system referred to in [22] as Systematic Lossy Error Protection (SLEP) in case of channel error rate increase allows to achieve graceful degradation of video quality without layered structure of the bit stream) and the absence of the prediction loop. The latter enables also codec-independent scalability that looks very appealing in a perspective of the growing variety of user terminals. Moreover, it allows the multiview coding by exploitation of inter-view correlation at the decoder side.

### 2.3.4 Future trends

The evolution of video coding technologies continues nowadays. Together with an availability of higher bit rate channels and an increase in achievable computational complexity that goes in line with Moore's law [71], the quality expectations, resolutions and demand for bandwidth are also rising. One of the main trends - to maximize the quality and minimize the bit rate - which was clearly notable in video compression development during the past decades - still remains central. The development of the HEVC standard continues the strategy of improving the rate-distortion performance due to the addition of new coding modes. The demand for increasing picture formats up to 4K ($3840 \times 2160$) and potentially up to 8K ($7680 \times 4320$) will be also satisfied by HEVC. As

the evolution of mobile video services continues towards the support of higher resolutions, upcoming standards need to be compliant with them as well. Already now a Call for Proposals for a scalable extension of HEVC has been announced [72] and the first works start to appear [73].

While performance of conventional coding improves, development of DVC algorithms continues as well. From the Slepian-Wolf and Wyner-Ziv theorems it follows that DVC performance potentially can be as good as the performance of conventional coding. Therefore, due to the existing performance gap between theoretical limits and practical results, there still remains a challenge of further improvement of the coding efficiency of DVC. However, even with the currently achieved coding efficiency the DVC framework shows its applicability to error-prone video transmission due to its error-resilience characteristics, and in wireless visual sensor networks due to distributed multiview video coding possibilities that allow to avoid communications between the cameras. This property opens some other perspectives for video surveillance and multiview video entertainment allowing low-complexity solutions. DVC and conventional approaches can also be combined in the future in order to achieve advances in rate-distortion performance.

Even if 3D video coding is out of the scope of this thesis, we consider it important to mention this research direction among future trends. The evolving interest in 3D video has caused an appearance of a variety of 3D video coding approaches like conventional stereoscopic and multiview video [21], mono/multiview video plus depth [74], layered depth video [75] and some others. The importance of 3D video has lead to the development of the Multiview Video Coding (MVC) standard [20, 21] that laid the basis for future work towards standardization of 3D and free-viewpoint television. A 3D Video Coding (3DVC) standard is currently under development and its goal is to define a unique format and compression technology for 3D video content. A new framework for scalable 3D based on wavelet or multiple description coding has also been proposed in several variations [76–79].

Within the development of video codecs it is also important to be able to effectively evaluate the quality of the reconstructed video sequences. There is no standardized approach for objective video quality assessment for sequences not having equal characteristics, and in this prospective the development of an efficient quality metric for comparison of the videos

with different temporal or spatial dimensions represents a challenging task. Even more challenging seems now the development of an approach for 3D objective quality assessment, where both 2D and 3D artifacts have to be taken into account.

## 2.4 Applications with constrained resources

An extensive development of wireless networks together with video compression systems has led to a new era of mobile wireless video communications. Mobile TV, video-on-demand, video streaming, online video gaming, videoconferencing and video surveillance are now widely used in everyday life. These systems are often subject to constraints such as memory, computational complexity, power, bit rate and maximum acceptable delay.

### 2.4.1 Bit rate constraints

Coding performance depends on the distortion and rate achieved by the codec. Bandwidth or bit rate constraints can also appear as communication plays a significant role in mobile video systems. In order to provide seamless communications, the available bandwidth should be comparable with the encoder output bit rate. Video source rate control algorithms can be utilized at the encoder side in order to minimize the source distortion provided that requirements for bit rate, power consumption and delay are satisfied.

### 2.4.2 Complexity and power constraints

Apart from bit rate constraints, handheld devices have other restrictions. The overall complexity of the encoder, consisting of the complexity of individual blocks and depending also on the needs for memory access, is often limited as well. In order to accelerate video encoder, fast algorithms (e.g. fast ME or fast Rate-Distortion Optimization (RDO)) can be implemented along with standard solutions requiring more heavy computations. Usually the encoder complexity has a direct connection to the functionality, i.e. more complex encoders have more available compression options and features. However, some properties can be added to originally light-weight encoders without making them significantly more

complex. In [6] we have described an approach for low-complexity scalable video coding based on JPEG-LS. As the solution is based on adding a second compression block based on the same JPEG-LS algorithm, this scheme can be efficiently implemented in hardware.

In emerging power-constrained systems like wireless video sensor networks, wireless video surveillance systems or wireless video endoscopes, the main focus is on the encoder power consumption. These systems with limited resources usually utilize a light-weight encoder. The recent advances in distributed source coding allow to use DVC approach in order to have a compression efficient and error resilient solution. In order to evaluate the complexity of DVC, we have compared the encoder power consumption of the state-of-the-art DVC codec and of the conventional approach based on the H.264/AVC standard [9].

In the situation in which only a software version of the compression algorithm is available, one of the ways of computational performance estimation is measurement of the encoding time. This is how H.264/AVC reference software [80] was compared to the state-of-the art DVC codec [31] in [81]. This approach has a disadvantage, as non optimized versions of software from the time consuming point of view lead to the incorrect comparison. Complexity can also be estimated in terms of basic operations, including addition, multiplication, shift and comparison. These operations can be further transformed into the power consumption values.

In our work we have demonstrated another way of evaluating power consumption. As the DVC codec and H.264/AVC use the same DCT and similar quantization techniques, we have proposed to compare only the power consumption of the remaining blocks of the schemes. By using power consumption/bit rate relation for CAVLC and Low-Density Parity-Check (LDPC) encoders used in H.264/AVC and DVC, respectively, and rate-distortion performance of these codecs, we estimated the relative power consumption for fixed PSNR values. Our results show that for a given quality the H.264/AVC compression in low-complexity configuration can achieve comparable coding performance and power consumption with the DVC solution.

### 2.4.3 Memory constraints

There are two reasons for memory restrictions: on the one hand, they appear due to practical implementation feasibility; on the other hand, higher system memory requirements can result in higher overall system costs that make a device less attractive commercially. Video compression systems using inter-prediction require memory storage of minimum one reference frame. For more complex systems with hierarchical GOP structure, full GOP needs to be stored in order to provide a complete reconstruction. In order to reduce the memory consumption when only Intra-coding is used, it is possible to divide the source frame into tiles that are compressed independently [7]. To achieve higher performance when tiling option is used, static region detection proposed in [10] can be used.

### 2.4.4 Delay constraints

Delay restrictions play a critical role in many applications (e.g. video conferencing or video surveillance) where the data needs to be assessed in real time. Generally, the overall system delay $\Delta T$ depends on many parameters: it consists of encoding processing delay $\Delta T_e$, encoder and transmitter buffer delays $\Delta T_{eb}$ and $\Delta T_{tb}$, channel delay $\Delta T_c$, receiver and decoder buffer delays $\Delta T_{rb}$ and $\Delta T_{db}$ and decoding processing delay $\Delta T_d$ (see Eq. 2.9).

$$\Delta T = \Delta T_e + \Delta T_{eb} + \Delta T_{tb} + \Delta T_c + \Delta T_{rb} + \Delta T_{db} + \Delta T_d. \tag{2.9}$$

It was shown [82] that if both encoder and decoder work in real-time, Eq. 2.10 holds true:

$$\Delta T_{eb} + \Delta T_c + \Delta T_{db} = L, \tag{2.10}$$

where $L$ is the accumulation time on the receiver side before the decoding and reproduction start, if the number of bits in the encoder buffer is less or equal to the effective buffer size [83].

An approach for low-latency video transmission using low-complexity compression approach has been described in [10].

## 2.5 Video quality evaluation for performance comparison

In Section 2.1.2 we have addressed the question of subjective and objective quality assessment and briefly described the way of obtaining subjective scores (MOS). Detailed recommendations for performing subjective tests can be found in [44]. An extensive overview of most of the main works on objective and subjective quality assessment for multidimensional scalability together with future trends and challenges is given in [84].

### 2.5.1 Objective assessment of video quality with different frame rates

As mentioned in Section 2.1.2, most of the well known objective quality models, such as PSNR, Mean Structural Similarity Index (MSSIM), Video Quality Metric (VQM), focus mainly on spatial artifacts leaving apart the question of the impact of temporal resolution on the visual quality.

Some subjective studies [34, 85–87] have analyzed the impact of different frame rates. These works lead to a similar conclusion: the impact of spatial distortion tends to be more crucial for the perceived quality than the impact of frame rate. An increase of importance of frame rate after acceptable SNR is achieved is reported in [88] and a high dependence of frame rate preference on the content is observed in [89, 90]. However, subjective tests require lots of resources and are not applicable to real-time quality monitoring. For these reasons researchers are trying to develop objective metrics that correlate as good as possible with the human perception of video sequences.

### 2.5.2 Existing objective metrics with focus on frame rate

Recently a few novel quality models with special focus on frame rate have been proposed [1, 25–27]. The metric [25] considering both quantization and frame rate consists of PSNR combined with an offset that depends on the frame rate and the level of motion in the sequence. In case of high frame rates, the metric is dominated by the quantization errors and is close to the PSNR, thus the offset is set to be small. Otherwise, the

PSNR value (measured on the temporally upsampled sequences in case of low frame rate) is compensated depending on the frame rate reduction and the motion speed. Feghali et al. reported higher correlation with subjective measurements than PSNR metric [25]. However, the compensation is not enough for very low frame rates, as in this case PSNR values for interpolated sequences do not vary significantly for different video qualities, which results in almost equal values of the quality metric at low frame rates.

Another quality metric - VQMTQ (Video Quality Metric considering Temporal resolution and Quantization) [26] - provides a good fit to the measured MOS scores. Ou et al. report better performance than the metrics proposed in other works with similar complexity [26]. The model uses sequence-dependent parameters; however, it is possible to predict these parameters based on the characteristics of the video such as frame difference, motion direction, and Gabor texture features.

Peng and Steinbach proposed a novel full-reference video quality metric called STVQM (Spatio-Temporal Video Quality Metric), based on PSNR, frame rate, and spatiotemporal activity measures [27]. Three CIF video sequences - "Mother & Daughter", "Foreman" and "Football" - with original frame rate of 30 fps were used in the experiments. Each sequence was temporally downsampled to 15, 10 and 7.5 fps and encoded with MPEG-4 to generate three different spatial quality levels. Results reported in [27] show that it performs significantly better than metric in [25] whereas the difference between VQMTQ and STVQM is not statistically significant. However, STVQM has some advantages, namely the use of only two standard video activity indicators that can easily be computed, compared to four parameters with significantly more complex interpretation used in VQMTQ. Moreover, STVQM has no codec-dependent parameters, unlike VQMTQ.

### 2.5.3 Subjective studies with focus on frame rate

The studies described above are lacking cross-validation on different subjective video quality databases, as the proposed models were trained and validated on ad-hoc subjective experiments. In [1] we proposed a quality metric that takes temporal resolution into consideration. The subjective results from [34] have been used both for training and validation of the

proposed metric. We have extended the validation part by including the subjective results by Ou et al. [91, 92] in our study.

Subjective data [91] has been obtained by 31 non-expert participants. After a training session (based on sequences "Soccer" and "Waterfall"), each user had to rate the quality of a shown video in a range from 0 (worst) to 100 (best). Four test sequences ("Akiyo", "City", "Crew", and "Football") of CIF resolution at original frame rate of 30 fps have been encoded with H.264/SVC (version JSVM 9.12) [93] in order to produce 4 temporal layers (with frame rates 30, 15, 7.5, 3.75) and 4 quality layers (corresponding to QP 28, 36, 40, and 44) using coarse grain scalability. A total of 64 processed video sequences have been used in the experiment.

Another subjective quality assessment test [92] has been performed on five CIF sequences ("Akiyo", "Foreman", "Football", "Ice", "Waterfall") at original frame rate of 30 fps. Similarly to the previous experiment, several temporal (frame rates of 30, 15, 7.5, 3.75) and quality (QP equal to 28, 36 and 40) layers have been obtained using JSVM 9.12 version of H.264/SVC. Sequences with varying frame rate were also included in this dataset, but we only use the data for sequences with constant frame rate. Subjective tests have been done by 33 participants who rated the viewed sequences from 0 to 100. The assessment has been performed following a protocol similar to Absolute Category Rating (ACR) [47].

**Obtaining quality scores from subjective preferences**

The two datasets described above provide subjective scores in terms of MOS. In [34] a new method for the evaluation of the relative impact of frame rate and spatial quality was proposed. Using a pairwise comparison method, the subjectively preferred path from the lowest quality level to the highest can be found. The way of converting the subjective preferences from pairwise comparisons into quality scores can be described as follows.

We assume that the impact of frame rate on perceived quality become negligible for a frame rate level equal or higher than 25 fps. Therefore, we consider that the quality of high frame rate sequences is defined by the spatial quality $Q$. This is why in order to evaluate $Q$ we can use any quality metric or index that is agnostic to the frame rate, for example PSNR or Structural Similarity Index (SSIM). There are $n$ test subjects who have reached position ($Q_x$, $FR_y$) on the quality plane, and their task

is to choose the preferred sequence between $A$: ($Q_{x+1}$, $FR_y$) and $B$: ($Q_x$, $FR_{y+1}$) (see Fig. 2.6). When $n_A$ subjects choose $A$ and $n_B = n - n_A$ choose $B$, transition probabilities $p_A$ and $p_B$ for a random person to choose $A$ or $B$, respectively, can be estimated:

$$p_A = \frac{n_A}{n}; \tag{2.11}$$

$$p_B = \frac{n_B}{n}. \tag{2.12}$$

Assuming that the perceived quality levels for $A$ and $B$ are reasonably close to each other, the perceived quality difference $\Delta PQ$ between $FR_y$ and $FR_{y+1}$ on the $Q$-scale can be estimated from the quality difference $\Delta Q$ between $Q_x$ and $Q_{x+1}$ by the following equation:

$$\Delta PQ = \Delta Q \frac{p_A}{p_B}. \tag{2.13}$$

Then, starting from the highest frame rate and highest spatial quality, we can compute values for $\Delta PQ$ cumulatively. If there are not enough test subjects who have occupied a certain node when traversing through the plane, the $PQ$ may have to be omitted for that node.

### 2.5.4 Proposed objective metric

The Perceptual PSNR (PPSNR) metric that uses conventional PSNR to model the spatial quality, proposed by us in [1] is computed as follows:

$$PPSNR = PSNR \cdot (1 - exp(1 - 10^8 \cdot FR \cdot PSNR^{-c})), \tag{2.14}$$

where $FR$ is the frame rate and $c$ is a content-dependent parameter. The value of $c$ is predicted from spatial and temporal characteristics of the video content. Experimentally obtained values of $c$ allow predicting its values for new contents by applying linear regression.

As reported in [1], the performance of the proposed metric on the subjective data from [34] is reasonably good compared to VQMTQ and STVQM. However, further experiments with other datasets revealed that the proposed single parameter model for computing PPSNR is not always sufficient to match with experimental subjective data. This is why we propose a new version of the model with two parameters, $a$ and $c$:

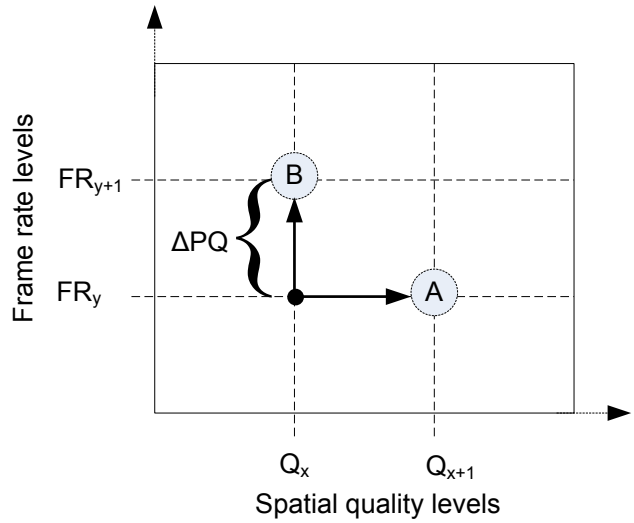



**Figure 2.6:** The choice of spatial quality or frame rate via pairwise comparison.

$$PPSNR = PSNR \cdot (1 - exp(a - 10^5 \cdot FR \cdot PSNR^c)). \tag{2.15}$$

In order to predict parameters $a$ and $c$, we have used slightly modified version of spatial and temporal activity indices $SA$ and $TA$ originally defined in [47]:

$$SA = mean_{time}(mean_{space}[Sobel(F_n)]), \tag{2.16}$$

$$TA = mean_{time}(std_{space}[F_n - F_{n-1}]). \tag{2.17}$$

In order to make the prediction of $a$ and $c$ more stable, we perform the prediction in the logarithmic space. Therefore, the following linear predictors are used:

$$a = -exp(\alpha_a + \beta_a \cdot SA + \gamma_a \cdot TA), \tag{2.18}$$

$$c = -exp(\alpha_c + \beta_c \cdot SA + \gamma_c \cdot TA). \tag{2.19}$$

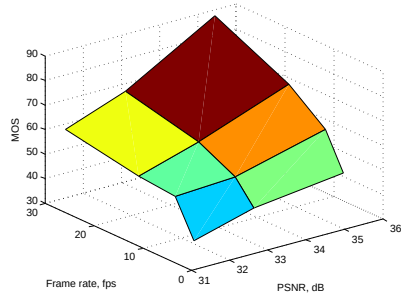
The prediction coefficients ($\alpha_a$, $\beta_a$, $\gamma_a$, $\alpha_c$, $\beta_c$, $\gamma_c$) are derived by first solving the optimal values of $a$ and $c$ for a training set with known subjective quality scores via nonlinear regression to match Eq. 2.15, and then solving the linear predictor coefficients in Eqs. 2.18 and 2.19 by applying linear regression to the solved values $ln(-a)$ and $ln(-c)$. The new metric has two major differences in comparison to the metric in [1]: we have introduced a new parameter $a$ in Eq. 2.15, and we have modified the predictor functions used in Eqs. 2.18 and 2.19. In addition, the definition of $SA$ had a minor change, since we used mean values from the Sobel filtered image instead of standard deviation.

### 2.5.5 Performance evaluation

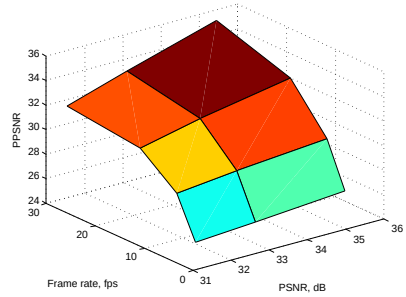
In our work [1] we exploited the database from the study [34] that provides subjective scores for five sequences. Ground truth quality scores were obtained from subjective preferences by the method described in Section 2.5.3. Such a small number of data did not allow us to create separate subsets for training and validation. In order to alleviate this, a linear regression was performed for each sequence separately using the "leave-one-out" method, i.e. parameters $\alpha$ and $\beta$ required for predicting $c$ were obtained using the known values of $c$ for the other four sequences. The major deficiency of this method of testing is the sensitivity to outliers, and therefore the performance varies highly across the sequences.

To make an extensive comparison of the modification of PPSNR metric described above, we have included datasets A [34], B [91] and C [92] in our evaluation experiments. Datasets B and C contain MOS scores, while dataset A provides subjective scores derived from pairwise comparisons and PSNR values. Therefore, we have been able to use different independently generated video databases for training, where we obtained the linear predictor coefficients $\alpha$, $\beta$ and $\gamma$, and validation.

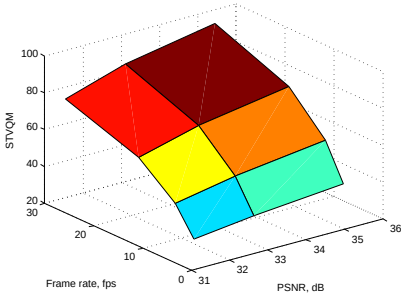
We have compared performance of our metric with VQMTQ and STVQM. Our results averaged over sequences in each individual dataset used for validation are shown in the Tables 2.1-2.3. The correlation between subjective scores and objective results obtained by metrics are demonstrated by Root Mean Square Error (RMSE), and Spearman and Pearson correlation coefficients. In order to average the values of correlation coefficients over the sequences in the dataset, we have used the



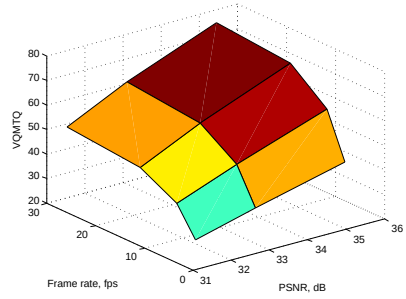

(a) MOS




(b) PPSNR
Spearman = 0.96, Pearson = 0.94




(c) STVQM
Spearman = 0.94, Pearson = 0.93




(d) VQMTQ
Spearman = 0.99, Pearson = 0.97

**Figure 2.7:** Relation between PSNR, frame rate and quality metrics

Fisher's Z-transform [94]. It is important to note that RMSE values are not comparable across datasets due to the difference of scales for dataset A and datasets B and C.

The provided results demonstrate that our metric PPSNR can compete with other objective metrics that take frame rate influence on the video quality into account. One advantage of the proposed metric in comparison to VQMTQ is the simplicity of parameters prediction. While VQMTQ parameters require complex computation from multiple characteristics of video content, PPSNR parameters only depend on spatial and temporal activity indices. VQMTQ shows good performance results on datasets B and C. However, it is necessary to note that these datasets come from the designers of VQMTQ.

As already reported in [1], metric performance depends on the content. However, results show also a dependency on the dataset used for training. For example, average correlation coefficients for dataset B are lower if PPSNR values were computed using $\alpha$, $\beta$ and $\gamma$ obtained on dataset A. Similar behavior can be noticed in the dataset C. On the other hand, PPSNR shows very good performance on cross-validated datasets B and C. These results support the hypothesis that predictor parameters for PPSNR metric are more reliable when obtained on the training dataset that uses the same subjective quality assessment method as the dataset for validation. However, reasonably good performance of PPSNR on dataset A using parameters from datasets B and C, and vice versa, confirms that the novel method for subjective quality evaluation proposed in [34] can be applied for assessment tasks along with MOS.

**Table 2.1:** Performance results for quality metrics for dataset A

| Metric | RMSE | Spearman | Pearson |
|---|---|---|---|
| VQMTQ | 1.40 | 0.93 | 0.94 |
| STVQM | 1.33 | 0.95 | 0.97 |
| PPSNR (trained on dataset B) | 1.58 | 0.94 | 0.94 |
| PPSNR (trained on dataset C) | 1.10 | 0.95 | 0.95 |

Fig. 2.7 shows the relation between PSNR, frame rate and the considered metrics for the "Ice" sequence from dataset C. PPSNR metric results were computed using dataset B as a training dataset.

In [1] we have used the "leave-one-out" method for validation of the

**Table 2.2:** Performance results for quality metrics for dataset B

| Metric | RMSE | Spearman | Pearson |
|---|---|---|---|
| VQMTQ | 2.81 | 0.98 | 0.99 |
| STVQM | 4.06 | 0.98 | 0.97 |
| PPSNR (trained on dataset C) | 3.72 | 0.98 | 0.98 |
| PPSNR (trained on dataset A) | 6.51 | 0.95 | 0.92 |

**Table 2.3:** Performance results for quality metrics for dataset C

| Metric | RMSE | Spearman | Pearson |
|---|---|---|---|
| VQMTQ | 3.69 | 0.97 | 0.96 |
| STVQM | 5.05 | 0.93 | 0.92 |
| PPSNR (trained on dataset B) | 4.18 | 0.94 | 0.94 |
| PPSNR (trained on dataset A) | 5.48 | 0.95 | 0.91 |

metric. These results are repeated in the Table 2.4.

**Table 2.4:** Euclidean distance between subjectively preferred paths and paths obtained by different metrics

| Metric | akiyo | city | coastguard | football | ice | average |
|---|---|---|---|---|---|---|
| VQMTQ | 0.65 | 0.34 | 0.27 | 1.24 | 0.37 | **0.57** |
| STVQM | 0.11 | 0.35 | 0.23 | 1.06 | 0.92 | **0.53** |
| PPSNR | 0.11 | 0.69 | 0.54 | 0.20 | 0.49 | **0.41** |

The new proposed metric shows very good results using this testing method as well. For example, for dataset C the "leave-one-out" method allows to achieve very good performance: the average Spearman and Pearson correlation coefficients are equal 0.94 and 0.96, respectively. PPSNR shows high correlation for this dataset when the parameters are defined using all sequences from the dataset for the prediction: 0.97 and 0.98 for Spearman and Pearson, respectively. These results demonstrate that PPSNR can have a very good performance if parameters $a$ and $c$ are predicted with high precision.

Performance results demonstrated above show that PPSNR can be used for comparison of video sequences with different frame rates. The simplicity of computation of content dependent parameters allows for

an easy utilization of this metric in any application that requires low-complexity. Cross-validation of the proposed metric on datasets provided by subjective tests performed in different research laboratories and in different countries confirms the versatility of the PPSNR metric.

## 2.6 Conclusion

In this chapter we have taken a look at low-complexity video compression without ME at the encoder side. Avoiding ME allows reducing the complexity and memory consumption of the encoder. However, in case of conventional video coding it leads to a loss of rate-distortion performance. Encoding in differential frame mode, though leading to an increase of required memory, allows achieving better compression results. A simple method of reducing temporal redundancy using detection of static regions that we proposed in [10] leads to a higher performance as well. Other approaches without ME at the encoder side, though yielding standard featured codecs in performance but applicable in resource-constrained scenarios, are based on MJPEG and DVC techniques.

In our work we discussed the two last named techniques and compared their performance with standard solutions. We have shown in [9] that an H.264/AVC encoder operating in differential frame coding mode can show comparable performance and complexity. It is not only the simplicity of the encoder that makes DVC an appealing solution: the scalability feature of DVC codec is useful in heterogeneous communication networks [95], visual sensor networks, and surveillance and streaming applications [96]. In the work [11] we have compared temporal scalability in DVC and H.264/AVC and concluded that the choice of one or another solution should depend on memory consumption requirements. In [8] we have proposed a scalable-to-lossless solution that achieves competitive performance with JPEG-LS, JPEG2000 and H.264/AVC in lossless mode. To the best of our knowledge, this is the only scalable-to-lossless solution based on DVC. Lossless compression using DVC can be a promising solution for the compression of hyperspectral images [97,98]; moreover, it can also be considered for other scientific and medical applications.

Scalable solutions play an important role in communication networks and surveillance applications where video data has to be delivered to het-

erogeneous clients with different capabilities. Scalability can be achieved not only with the DVC approach: JPEG2000 and H.264/SVC standards produce scalable bit streams and allow image and video decoding in various spatial, quality and temporal resolutions. However, a drawback of these solutions compared to DVC is the complexity of the encoder, which limits its usage in applications with complexity and power consumption constraints. We have proposed two simple scalable (progressive) solutions suitable for image and video coding based on JPEG and JPEG-LS [6, 7]. Performance results show that our low-complexity solutions can compete with standard approaches. Moreover, our scalable versions of JPEG and JPEG-LS can also be combined with UEP, which allows error-resilient wireless transmission over error-prone channels. This and other previously named advantages of scalable compression show the potential of scalable video coding. The planned scalable extension of HEVC [72], which, however, does not refer to a low-complexity solution, confirms the importance of scalable approaches.

There still remains an open question with the comparison of video data with different temporal and spatial resolutions obtained by scalable video codecs. Few quality metrics taking these factors into account have been proposed [25–27], however they do not show a perfect correlation with subjective opinion scores. We have proposed a novel quality metric for the comparison of sequences with different frame rates [1]. Results of the performance evaluation show that on average it allows achieving slightly better performance than other quality metrics that consider the temporal dimension. We described its improved version in Section 2.5.4 and demonstrated cross-validated results on different datasets.

Many ongoing research works show that video compression evolution still continues. It is supported by the development of related science fields like space, medical, geographic that use video and image data. A growing popularity of video entertainment applications also makes its contribution to the development of video communications. As new use cases and systems that deal with video data appear, new methods and approaches for video compression have to be developed as well. These methods often have to take some system restrictions into account, which makes the development of efficient methods and algorithms even more challenging.

# Chapter 3

# Video transmission systems

## 3.1 Introduction

Within just a few decades wireless technologies have evolved from the first generation (1G) to the fourth (4G), providing nowadays mobile broadband access to handheld devices and offering a large amount of services. A wireless network can be categorized by its scale in Wireless Personal Area Network (WPAN), Wireless Local Area Network (WLAN), Wireless Metropolitan Area Network (WMAN), Wireless Wide Area Network (WWAN). These networks are typically based on different transmission technologies, therefore they have different properties and provide different services. Wireless video transmission is possible over all these types of networks but in order to have an efficient communication system, their characteristics need to be taken into account while choosing an optimal video compression scheme.

Wireless video transmission provides a variety of services like Mobile TV, video-on-demand, video conferencing or online video gaming. In general these applications can be divided into two types: "conversational" (e.g. video telephony) and "non-conversational" (e.g. streaming/broadcasting). The main difference between them is in delay constraints which are much more strict for real time data transmission ("conversational" applications).

Video communication system can be described by the joint video compression (see Fig. 2.1) and transmission scheme shown in Fig. 3.1. The source data is sent to the encoder that can be combined with the

rate control unit in order to satisfy the requirements for average delay or average bit rate by choosing appropriate parameters for compression. The channel feedback provides information to the rate control unit about the channel state. A transmitter buffer [99] is used for constant bit rate transmission over communication channel. At the receiver side a similar buffer is used for accumulation of the compressed data before the decoding process starts. The decoder's output is a reconstructed sequence with compression distortion in case of lossy coding. As this scheme involves compression and transmission, end-to-end distortion can be induced both by quantization and channel. End-to-end delay is defined by the time between the instant, when coding block has been sent to the encoder, and the instant, when this block has been reconstructed at the receiver side.

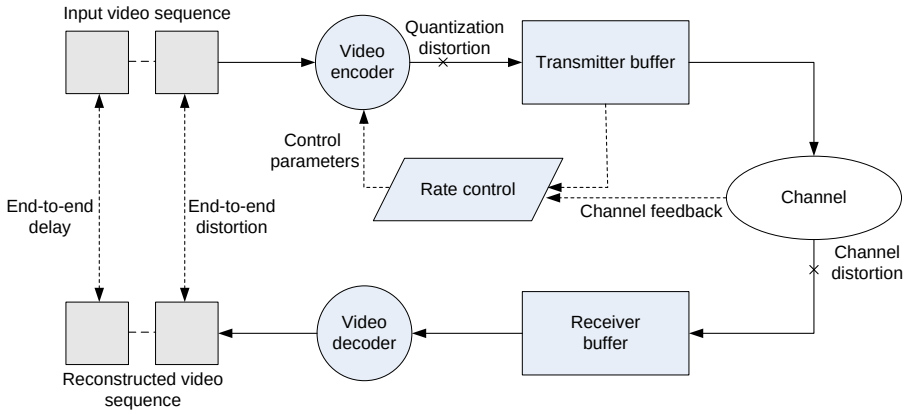



**Figure 3.1:** Video transmission scheme

Wireless video communication systems are a good example of application for video compression algorithms. In order to achieve optimal performance such systems require cross-layer optimization techniques that consider both compression and transmission aspects. Each video communication system requires an individual approach as optimization techniques may differ due to the diverse characteristics of both chosen compression algorithm and transmission network. The growing popularity of high-speed WPANs and ubiquity of mobile networks make researchers looking for efficient techniques for video transmission over such

types of networks. The development of such techniques becomes more difficult due to the constrained resources of client terminals, e.g. handheld devices, that pose additional restrictions to video compression and transmission systems.

## 3.2 Modern communication systems

### 3.2.1 Network types

Video transmission can be performed through various types of wired and wireless networks. The most commonly used technology for wired Local Area Network (LAN) is Ethernet [100]. If information is transmitted between devices without wires, such networks can be distinguished based on the operation area. Local areas like houses, cafes, university campuses are often covered with WLANs, e.g. based on the IEEE 802.11 standard [101]. Such systems usually operate in infrastructure mode where an access point is required for the connection to the internet, but wireless communications can be also used to establish ad-hoc networks within several user devices. For shorter range communications (up to 10 meters) WPANs are used. For example, such networks can be based on Infrared [102] or Bluetooth technologies [103]. There is also a low-rate WPAN standard [104] that laid the basis for the ZigBee specification [105] suitable for wireless sensor networks, and a group of high-speed WPANs based on 60 GHz. In a metropolitan area, it is more appropriate to talk about WMANs that allow to connect different buildings. WiMAX network based on IEEE 802.16 standard [106] is a good example of WMAN. The WiMAX technology represents an alternative solution to classical wired approaches using copper cables or fiber optics. Finally, WWANs have worldwide coverage and provide mobile access to the Internet from any area that has a signal coverage. Mobile wireless networks starting from the second generation (2G) can be referred to as WWAN.

### 3.2.2 Mobile generations

The transition from analog (1G) to digital (2G) transmission has happened in the 1990s. At that time cellular phones provided short messaging and low speed data services. The capacity of 2G cellular systems has

improved in comparison to 1G systems, but further demand for greater bandwidth resulted in the development of the third generation mobile telecommunications (3G), providing broadband access to mobile phones and handheld devices. In order to be called "3G", the technology needs to meet the International Mobile Telephone 2000 (IMT-2000) project requirements, providing 144 kbps of throughput at mobile speeds (in a car or a train), 384 kbps at pedestrian speeds, and 2 Mbps in indoor environments [107]. The upcoming 4G techology (official name "IMT-Advanced") will provide mobile ultra-broadband Internet access, wide range of data rates and capabilities for high-quality multimedia applications with higher performance and quality of service like HD Mobile TV and 3D television. The targets for peak data rates are 100 Mbit/s for high mobility (mobile access) and 1 Gbit/s for low mobility (local access) [108, 109]. Two technologies under development - LTE Advanced (Long-Term-Evolution Advanced) and IEEE 802.16m or WirelessMAN-Advanced - are the candidates for the evolving 4G standard. However, none of the currently implemented systems does comply fully with the IMT-Advanced requirements.

Nowadays, most of the mobile phones are working in GSM, WiMAX and 3G networks. The last named is becoming ubiquitous. As wireless communication technology advances, 3G networks provide global communications with various services including telephony, messaging (including multimedia), high data rate internet access, videoconferencing, global positioning, high quality music and video downloading, and online gaming capabilities.

### 3.2.3 Wireless personal area networks

In the recent years high-speed WPANs are also becoming more and more popular. They have low power data transmitters that allow long battery lifetime and provide scalability in terms of bit rate - from low to very high (20 Mbps - 28 Gbps). They are mostly used for short-range communications between personal devices and have a high security level due the fact that coverage area is limited to room space.

As the frequency range around 60 GHz (from 4 to 9 GHz within 57-66 GHz) has been regulated for unlicensed use worldwide, new standards based on 60 GHz started to appear. In 2009 WPAN Standard 802.15.3-

2003 was extended with an alternative physical layer (PHY) [28]. The first consumer application of 60 GHz technology - WirelessHD [29] - provides wireless connection and data exchange between a wide range of devices, including laptops, televisions, Blu-ray players, gaming consoles etc. The WiGig specification [110] utilizes the 60 GHz band to provide data rates up to 7 Gbps. Though based on the 802.11 standard, it has much more spectrum available than the 2.4 GHz and 5 GHz bands utilized by Wi-Fi products, and therefore supports faster transmission rates.

Due to high data rates, these 60 GHz networks can transmit High-Definition (HD) video between devices over wireless channels instead of wired cables for high quality video services such as HD video conferencing and distributed video gaming. Even though uncompressed data transmission is possible [111, 112], video compression is preferable as it allows to decrease the overall power consumption for the transmission [113].

These two types of networks described above - 3G and WPANs - are currently the most relevant for wireless video transmission due to the availability of high data rates. The development of novel solutions or optimizations of existing schemes for video transmission over such networks still remains actual. This task becomes even more complex, as many relevant applications have in addition power consumption and delay constraints.

## 3.3 Systems with constrained resources

### 3.3.1 Delay constraints

Delay is critical for real-time applications, e.g. conversational services like video conferencing or online video gaming. The maximum tolerable end-to-end delay for such kind of applications is estimated around 125-250 ms [114, 115].

As described in the Section 2.4, the end-to-end delay in the system depends on the particular delays during encoding and transmission processes (see Eq. 2.9).

Equation 2.10 shows that the accumulation time on the receiver side before the decoding and reproduction start is equal to the sum of encoder buffer delay, channel delay and decoder buffer delay, if the number of bits

in the encoder buffer is less than or equal to the effective buffer size.

In order to calculate the effective buffer size $b_{eff}(t)$ [83] for video transmission at the moment $t$, it is necessary to know the future channel throughputs $c_i$:

$$b_{eff}(t) = \sum_{i=t+1}^{t+L\cdot f\cdot N} c_i, \tag{3.1}$$

where $L$ is the required maximum latency, $N$ is the number of coding units in the frame and $f$ is the frame rate.

It is not possible to calculate the exact size of the effective buffer for time varying wireless channels, as future channel rates cannot be known. In this case, an estimation $\hat{b}_{eff}(t)$ obtained with a channel model [116] can be used, which, however, does not guarantee that at any instant $t$, $\hat{b}_{eff}(t) \leq b_{eff}(t)$. If $\hat{b}_{eff}(t)$ exceeds $b_{eff}(t)$, the latency requirements (2.10) do not hold true. To restore the required latency values it is possible to apply an approach that we proposed in [10].

In 3G systems, buffering latency can influence efficient power management as well. As was shown in [2], in order to comply with buffering latency restrictions, a mobile device has to stay in the state with low power consumption for a shorter time, which leads to an increase of the overall power consumption.

### 3.3.2 Power constraints

The total power consumption for video communication systems consists of power consumption needed for compression and the one needed for data transmission. Transmission power consumption depends on the amount of data that needs to be transmitted, that in turn is related to the efficiency of the compression. However, compression performance is also directly connected to the power consumption: better performance can be often achieved due to the increase of computational complexity of the algorithms and thus the increase of the power consumption for compression. An illustration of power consumption dependency on the encoding complexity parameter $CP$ is shown in Fig. 3.2. The solution between the conflicting issues results in achieving a computation-communication trade-off that for video communication systems is often called the Power-Rate-Distortion (P-R-D) trade-off. This trade-off can be applied to the

downlink communication as well, where the computation part refers to the power consumption of the decoder.

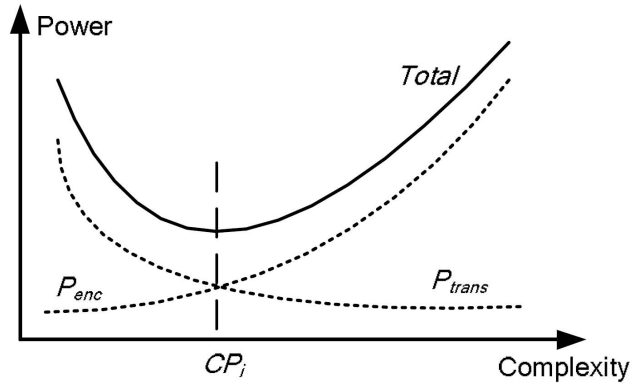



**Figure 3.2:** Relation between power and complexity [117]

In handheld devices, where power remains one of the most critical resources, it is essential to have an efficient power management. There are many ways of performing it on a mobile device. Such devices consist of several blocks like CPU, memory, network interfaces etc., which require different power consumption strategies to be applied on [118–120].

In most cases, power management techniques focus on the device power consumption. It is obvious that here some fundamental trade-offs like spectrum efficiency - energy efficiency, bandwidth - power, delay - power and quality - power appear [121]. In our analysis of the power consumption of the 3G transition state machine [2], we consider the two last named trade-offs. We propose a method for parameters selection for the 3G transition state machine that allows to decrease device power consumption. We also describe the cases with several restrictions like signaling traffic, buffer size and latency. As we apply this technique for uplink video transmission over 3G networks, we show the potential of content-dependent power management. For example, a mobile device can balance between video quality and transmission power. By defining the desired quality range, it is possible to avoid using additional power for unnecessary improvement of the video quality.

Power management is possible in the downlink mobile communications as well. It has been noted [119, 122–124] that the power consump-

tion of the receiver can be lowered if the transmission occurs in bursts. The main idea of such approach implemented in DVB-H [125] is that the receiver operates for a short time interval, during which the part of the video data is received. Then, the receiver turns off the radio parts completely while the video is being constantly decoded. The use of scalable codecs in this solution allows meeting the trade-off between power consumption and video quality in handheld receivers [51, 126].

## 3.4 System optimization

Efficient management of video communication systems requires cross-layer optimization techniques. Optimization methods can be focused e.g. on the end-to-end delay, power consumption and complexity. Cross-layer communication allows the system to use the information about operation on other layers. Such information like remaining battery capacity or channel conditions is useful to adjust the system parameters in order to achieve optimal performance.

### 3.4.1 Optimization for fiber-wireless transmission

The ideal compression system for wireless transmission over high-speed networks has not been developed yet. In order to perform video transmission over such networks, an optimization of the existing solutions should be done.

In [3] we have demonstrated the HD video distribution over fiber-wireless networks. For such a setup that can be used for both conferencing applications and distributed video gaming, some requirements need to be satisfied. These applications have delay and power consumption constraints, plus they require error-resilient transmission. In order to fit a realistic scenario, we have chosen the H.264/AVC codec that covers a wide range of wireless applications [114]. To comply with the requirement for low delay, the Intra coding mode was chosen. It allows avoiding long buffering time as reference frames are not used for prediction. In addition, the elimination of inter-prediction reduces the complexity of the algorithm. Instead of Context-Adaptive Binary Arithmetic Coding (CABAC), a lower complexity solution for entropy coding - Universal Variable-Length Coding (UVLC) - was used [64].

Intra mode allows at the same time to improve the error resilience together with other optional tools: slicing, data partitioning, Flexible Macroblock Ordering (FMO), Arbitrary Slice Ordering (ASO), redundant coded slices. In [3] two of these tools - FMO and slicing - have been used. The results have shown that employed tools allowed to increase the robustness of video for fiber-wireless transmission over 60 GHz channel.

### 3.4.2 Rate-Distortion-Complexity optimization

In Section 3.3.2 we have mentioned that complexity/power consumption for the whole video communication system depends on its two components - compression and transmission. However, these components influence each other: higher compression efficiency requires higher complexity and therefore, higher power consumption, but at the same time it allows to reduce bit rate and save on transmission power, and vice versa (see Fig. 3.2).

One of the goals of compression is to optimize the quality of the reconstructed video data. The limited bandwidth of the communication channels put additional constraints on the bit rate. An optimal compression solution under constrained bit rate can be found by Rate-Distortion Optimization (RDO). In mobile video communications power and computational complexity are two other constrained resources, therefore, in order to find an optimal solution under all constrained resources, it is not enough to perform only RDO. In this case, it is necessary to perform Resource-Distortion optimization [127], which includes P-R-D and Rate-Distortion-Complexity (R-D-C) analysis and optimization [128–132]. R-D-C optimization is part of the complexity management system, which allows data encoding within a certain complexity limit, i.e. complexity budget. An optimization can be done offline, i.e. already at the level of system design, or online, i.e. during data processing, allowing to adapt the algorithm to the current conditions.

Complexity limitations are caused by the device hardware capabilities, e.g. the operating speed of the processor and its working states (idle levels, etc.). This implies that there is a hardware limit for computational resources on the mobile device. Power restrictions are usually connected to the battery capacity, which is also limited on mobile devices.

Experimental studies show that the contribution of the encoder to

the overall power consumption for video communication systems is high and it usually exceeds the contribution of transmission power [133, 134]. Encoder power consumption depends on the complexity of the compression algorithm. Therefore, R-D-C optimization for the encoder can play a significant role in the complexity management system. Usually R-D-C optimization is based on efficient bit allocation in order to maximize the quality under complexity constraints. Some approaches for R-D-C optimization have been designed for the H.264/AVC standard [117, 132, 135]. They are typically based on optimization of Motion Estimation (ME), as it represents one of the most resource-consuming units of the codec, and controlling the mode decision operations.

**Rate-Distortion-Complexity optimization for HEVC**

We extended the idea of restricting the amount of coding modes tested during RDO to the upcoming High Efficiency Video Coding (HEVC) standard. The novel CU data structure of HEVC, described in Section 2.3.2, allows to control the complexity by varying the partition depth. One approach based on dynamic adjustment of the partition depth is described in [136]. In this work CU depth, which defines the smallest block size that can be used for encoding, is selectively constrained by a complexity control algorithm in order to comply with the complexity budget. Their idea is based on the assumption that co-located areas in adjacent frames are more likely to have similar behavior and consequently similar values of maximum depth; therefore it is possible to restrict depth values in a chosen number of frames basing on the maximum depth values in previous frames (in corresponding areas). This method allows achieving good precision in fulfilling target complexity and allows a wide range of complexity reductions.

We propose our solution for the task of complexity management for HEVC. In contrast to the approach described above, we find depth values for each Group Of Frames (GOF) in the sequence in the way that rate-distortion performance is close to optimal given the complexity budget. This allows us not only decreasing the depth values, but also increasing them for particular frames if it appears to be beneficial from a R-D-C point of view. HEVC does not allow changing the depth while a sequence is being encoded. This means that for a chosen encoder configuration, four complexity levels are available at maximum (referred

to four possible partition depth values). In our approach we encode each GOF independently. Using predictive techniques and game theory methods, we can define depth values independently for each encoded GOF, which allows us to control the complexity and perform the encoding of the sequence within a given complexity budget.

The proposed algorithm works in the following manner. After encoding the first GOF with all possible depth values and obtaining distortion, bit rate and complexity characteristics (measured as performance time), we can get the parameters $\alpha$ and $\beta$ separately for I and P frames for each depth value $d$ from the following dependencies:

$$Complexity = \alpha_1^d \cdot SAD + \beta_1^d, \tag{3.2}$$

$$Rate = \alpha_2^d \cdot SAD + \beta_2^d, \tag{3.3}$$

$$Distortion = \alpha_3^d \cdot SAD + \beta_3^d. \tag{3.4}$$

These functions allow us to predict complexity, rate and distortion values for the following GOF knowing the SAD values of its frames. The values of $\alpha^d$ and $\beta^d$ are updated to the values $\alpha'^d$ and $\beta'^d$ after encoding the new GOF with the chosen depth value $d$. Prediction for other depth values $d^*$ is done using previous values of $\alpha^{d*}$ and $\beta^{d*}$. Then, the results of the prediction are proportionally scaled in the following manner:

$$Complexity'(\alpha^{d*}, \beta^{d*}) = Complexity(\alpha^{d*}, \beta^{d*}) \cdot \frac{Complexity(\alpha'^d, \beta'^d)}{Complexity(\alpha^d, \beta^d)}, \tag{3.5}$$

$$Rate'(\alpha^{d*}, \beta^{d*}) = Rate(\alpha^{d*}, \beta^{d*}) \cdot \frac{Rate(\alpha'^d, \beta'^d)}{Rate(\alpha^d, \beta^d)}, \tag{3.6}$$

$$Distortion'(\alpha^{d*}, \beta^{d*}) = Distortion(\alpha^{d*}, \beta^{d*}) \cdot \frac{Distortion(\alpha'^d, \beta'^d)}{Distortion(\alpha^d, \beta^d)}. \tag{3.7}$$

Predicted complexity, rate and distortion values are used later to define the optimal values of the partition depth for the following GOFs using a game theory approach.

### Game theory

Game theory [137] is used in many science fields, including economics, social, political, and computer sciences. It is a common approach to solve the optimization problem of resource allocation [138, 139].

A game consists of the interaction of $n$ players that take decisions in order to maximize their own utility. The decision chosen by the $i$-th player at the time moment $t$ is referred to as a strategy and is represented as $\sigma_{i,t}$. The set of strategies chosen by $n$ players at the moment $t$ can be defined as $\sigma = [\sigma_{1,t}, \sigma_{2,t}, ..., \sigma_{n,t}]$.

The utility or payoff function $u(\sigma)$ depends on the array of strategies $\sigma$. The utility function $u_i(\sigma)$ for player $i$ is defined so that

$$\begin{cases} u_i : C_1 \times ... \times C_n \mapsto \Re \\ u_i : \sigma \to u_i(\sigma) \end{cases} .$$

A utility function is a criterion that allows to compare different strategies, i.e. the set of strategies $\sigma_1$ is better than the set of strategies $\sigma_2$ if $u(\sigma_1) > u(\sigma_2)$.

A game with $n$ players where the $i$-th player chooses its strategies $\sigma_i$ in the set $C_i$ and has a utility function $u_i(\sigma)$ defined above, can be expressed in strategic form: $(C_1, ..., C_n, u_1(\sigma), ..., u_n(\sigma))$. For a game $G = (C_1, ..., C_n, u_1, ..., u_n)$ a given configuration $(\sigma_1^*, ..., \sigma_n^*)$ is a Nash equilibrium if the strategy $\sigma_i^*$ is the best response of player $i$ to the strategies $(\sigma_1^*, ..., \sigma_{i-1}^*, \sigma_{i+1}^*, ..., \sigma_n^*)$, i.e.

$$u_i(\sigma_1^*, ..., \sigma_{i-1}^*, \sigma_i^*, \sigma_{i+1}^*, ..., \sigma_n^*) > u_i(\sigma_1^*, ..., \sigma_{i-1}^*, \sigma_i, \sigma_{i+1}^*, ..., \sigma_n^*)$$

for $\forall i \in N$ and $\forall \sigma_i \in C_i$.

In our problem statement, a game consists in the interaction of $n$ players (GOFs) that choose the configuration, i.e. the value of the partition depth, in order to maximize their own utility $u$. Each player $i$ has a choice between $m$ values of partition depth, i.e. the overall amount of possible configurations is equal to $P = m^n$. For each configuration $p \in P$ we find a utility function $u(p)$ that takes into consideration both predicted bit rate and complexity of the following GOF. Afterwards we search for the configurations that provide Nash equilibria. The configuration that gives the best payoff according to the prediction will be a solution of the optimization task for the following GOF. Afterwards

the GOF is encoded using the found depth value. Then, the complexity budget for the remaining GOFs is updated using the actual encoding complexity of the encoded GOF. Finally, rate, distortion and complexity characteristics for each depth of the next GOF are predicted, and the game theory approach is applied again. These operations repeat until the whole sequence is encoded.

The game theory approach described above can be extended to a bigger amount of available configurations. For example, each player can choose between the configurations that depend not only on the partition depth, but also on the quantization parameter and some other configuration options that affect the R-D-C performance of the algorithm.

### Performance results

We have performed R-D-C optimization on three CIF video sequences "Crew", "Mobile" and "News" that represent three different kinds of contents: high motion, texture and low motion. 70 frames per sequence were considered in our experiments. Each sequence was encoded as GOFs, consisting of 1 I frame and 4 P frames. The chosen CU size was $64 \times 64$.

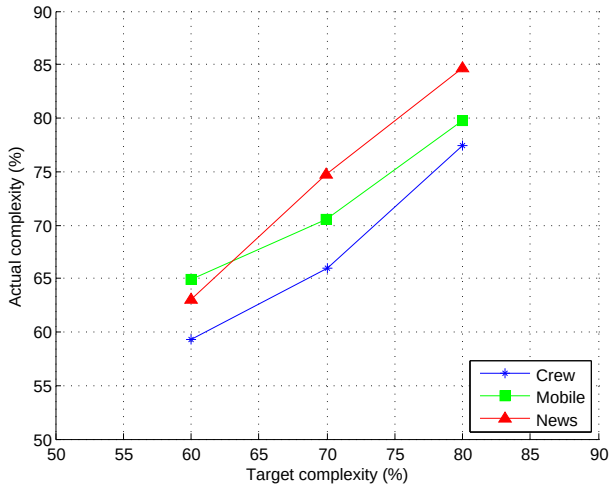



**Figure 3.3:** Relationship between the target and actual complexity

As it can be seen from Fig. 3.3, the proposed game theory approach

combined with prediction provides a quite accurate control of the complexity. The maximum deviation from the target complexity is around 5%. The complexity is measured by processing time, and the actual complexity refers to the time spent for encoding of the sequence de facto.

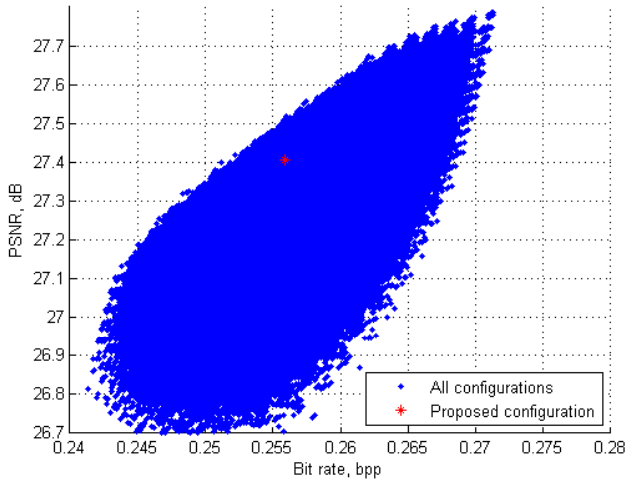
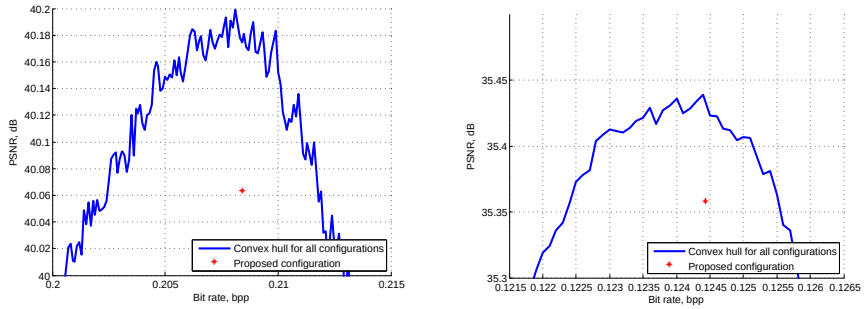



**Figure 3.4:** Performance of different configurations for the sequence "Mobile", 70% complexity level

Figure 3.4 shows all the configurations for the sequence "Mobile" that satisfy the defined complexity level of 70%. As it can be noticed, the configuration found by our method is quite close to the optimal ones providing the best rate-distortion performance.

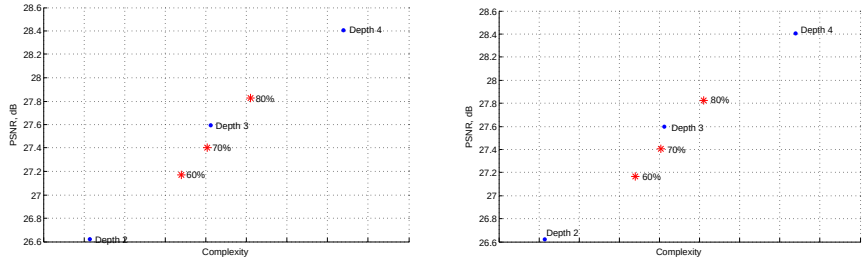
Figure 3.5 shows the performance of the proposed method in comparison to the optimal solutions found by offline optimization. It means that complexity (performance time) values for all possible configurations have been stored in advance and optimal solutions satisfying the defined complexity level have been found by full search. Table 3.1 shows the PSNR difference in dB between proposed and optimal configurations, compared at equal bit rates.

Figure 3.6 demonstrates that the proposed R-D-C optimization provides flexibility for the HEVC encoder in terms of complexity. As already mentioned, the current HEVC encoder allows 4 complexity levels at maximum. These levels are referred as Depth 2, 3 and 4 in Fig. 3.6. Depth 1 is not allowed for CIF sequences with CU size equal $64 \times 64$. Our solu-




(a) "News", 60% complexity level (b) "Crew", 80% complexity level

**Figure 3.5:** Performance results for the proposed solution compared to offline optimization




(a) "Crew" (b) "Mobile"

**Figure 3.6:** Flexibility of the HEVC provided by R-D-C optimization

**Table 3.1:** PSNR difference (dB) between the proposed and optimal configurations

| Sequence | 60% | 70% | 80% |
|---|---|---|---|
| Crew | 0.13 | 0.18 | 0.08 |
| Mobile | 0.13 | 0.11 | 0.10 |
| News | 0.10 | 0.16 | 0.12 |

tion allows HEVC performance with the complexity level defined by the end user. We have shown the performance in terms of PSNR/complexity relationship, as our setup implies the use of a rate control algorithm that allows producing bit streams with similar bit rates. However, slight deviations from the target bit rates are possible, especially on the short sequences like the ones used in our experiments (70 frames).

Our results for RDC optimization for HEVC (see also Appendix B) evidence a small PSNR loss against the solution that can be found by offline optimization. In contrast to the work [136], we do not only decrease the complexity of HEVC, but also provide a solution that gives rate-distortion performance close to the optimal one. The use of rate control for HEVC allows performing optimization under bit rate and complexity constraints at the same time.

Our approach shows the possibility of controlling the complexity of HEVC allowing its use in power-constrained devices. Optimization of the power consumption of the device through the complexity control of the encoder implies accessing the state of the hardware resources in the device. This means that R-D-C control should be performed from a cross-layer optimization point of view. This optimization can include another layer - transmission - if compressed video data has to be sent over the network. As described in the Section 3.3.2, such cross-layer optimization will result in finding computation-communication trade-offs.

## 3.5 Conclusion

Modern communication systems have been steadily evolving in the last years so that just in few decades we have passed from 1G to 4G networks. Digital transmission significantly advanced the development of wireless communication systems, so that now we have high data rates and capabilities for various applications. Networks availability and their variety,

convergence of wired and wireless networks, ubiquity of 3G mobile systems, popularity of social networks and conversational applications - all this has contributed to the development of the concept of "anytime anywhere" wireless connectivity.

However, wireless connections require additional power consumption from the client terminal. Mobile communication systems often include handheld devices that have additional constraints like power consumption. In order to prolong battery lifetime, it is possible to perform complexity management. We have shown in [2] that some modifications in the 3GPP transition state machine can bring significant benefits in terms of device power consumption. Apart from the ideal case, we have provided the analysis of system performance in case additional restrictions - like latency, amount of signaling traffic or buffer size - appear. As the share of video in mobile data traffic is continuously growing, we have discussed video data transmission over 3G networks and shown power-quality trade-offs that allow to perform adaptive power control on the mobile device.

Data transmission can be more efficient if a rate control system that helps to maximize the quality under the bit rate constraints is embedded. We have presented a low-latency low-power video source rate control and provided practical results for video transmission over high-speed WPANs using the proposed rate control [10]. If rate control is combined with power/complexity control, we can speak about P-R-D or R-D-C optimization. As power consumption of mobile communication systems consists both of compression and transmission power that depend on each other, a trade-off between these conflicting issues needs to be found. However, higher power consumption in video communication systems refers to encoding procedures than to transmission. Power management systems for encoders similar to the one presented in Section 3.4.2 allow to achieve another trade-off - between power consumption/complexity and resulting video quality, so that efficient but complex video coding schemes like HEVC can be applied to power-constrained devices. If all aspects including rate-distortion performance, power consumption and delay are taken into account, we can speak about Resource-Distortion control that allows to achieve efficient transmission.

Thus, many applications that deal with video data to be transmitted over wireless networks are becoming more and more popular. With high

bandwidth available in some types of the networks like 60 GHz WPANs, it is possible to transmit this data even in uncompressed form, but efficient video compression can bring additional advantages apart from the bit rate decrease. Video encoding with small compression ratios can lead to overall power savings in the system. In the perspective of growing resolutions up to $4K$ and even $8K$, video compression seems to be essential. Moreover, scalable bit stream allows the efficient usage of Unequal Error Protection (UEP) to provide error resilient transmission. However, error resilient transmission can be achieved by employing different tools. We have experimentally demonstrated 60 GHz Radio over Fiber (RoF) transmission of compressed HD video and the effect of error-resilience tools in H.264/AVC for 60 GHz RoF setup [3]. In particular, we have shown that these tools improve robustness against impairments that occur in 60 GHz fiber-wireless channel.

Bandwidth demand leads to the development of new generations of wireless networks and these networks require new approaches to handling of the video data. Thus, advances in communications and growing share of video in overall data traffic pose new challenges for the research in video coding as well.

# Chapter 4

# Conclusion and Outlook

The fast development of video communications that has taken place in the last decades has lead to many changes in our everyday life. Video chats, Mobile TV, wireless video surveillance, wireless home theaters - all these applications have entered in our life and almost immediately took very strong positions so that nowadays it is hard to imagine life without them. Such a variety of applications leads to a variety of technological approaches used in them. Low and high complexity compression algorithms, real time and non real time systems, scalable and multiview video codecs - just to name a few of them. These systems are often subject to memory, complexity, bit rate or other constraints that pose additional limitations to the compression systems. As compressed video data is often sent over communication systems, additional challenges appear. Transmission systems add further restrictions to power, bandwidth and delay, so an optimal solution has to be found taking all aspects of the joint compression and transmission system into account. Greater demand for bandwidth leads to new generations of wireless networks, higher resolutions lead to the evolution of compression standards so the development of video mobile communications continues and new approaches need to be found.

This thesis covered several important aspects regarding low-complexity wireless transmission of compressed video under constrained resources. We have proposed low-complexity scalable (progressive) solutions for video compression based on JPEG and JPEG-LS image coding standards. These two schemes can be applied in low-power systems where

in addition low memory consumption is required, and our results show that in such constrained scenario our methods compete with standard solutions.

We have studied scalability aspects in the novel Distributed Video Coding (DVC) paradigm as well. DVC is a low-complexity approach that together with the scalability and other features seems a promising solution for many applications. In our work we have compared DVC and H.264/AVC in terms of encoder power consumption and temporal scalability. Our results show that both solutions can provide satisfactory results for low-complexity coding and the choice between them should be made based on other restrictions and requirements of the system. We have also proposed a scalable-to-lossless DVC codec that outperforms other lossless solutions, i.e. JPEG-LS, JPEG2000, H.264/AVC, allowing around 5 - 13 % bit rate savings.

In continuation of the scalability topic, we have proposed a novel metric for objective quality assessment for video coding with temporal scalability. Such metrics are essential for comparison of video sequences with different frame rates produced by scalable video codecs. Our metric shows a high correlation with the subjective scores. We have also demonstrated the applicability of the novel method for subjective quality evaluation proposed in [34] to the tasks of quality assessment along with Mean Opinion Score (MOS).

Orienting on low-complexity solutions, we have performed power consumption analysis of compression and transmission systems with a focus on the power-distortion trade-off. We have proposed a power consumption model for the most consuming state of the Radio Resource Control (RRC) in 3G. Our model provides a good approximation to the experimental results. We have proposed an optimization for the 3GPP transition state machine that allows to decrease device power consumption. We have also discussed constant bit rate video transmission over 3G networks and shown the benefits of the proposed method in terms of quality vs. power consumption.

We have discussed low-delay and low-power video transmission over Wireless Personal Area Networks (WPANs). This included the optimization for joint fiber-wireless transmission over 60 GHz, the demonstration of error-resilience tools performance in fiber-wireless link, and a proposal of a video source rate control for low-latency video transmission.

Performance results for rate control, providing lossless quality in "good" channel states and guaranteeing acceptable quality for a given throughput in "bad" channel states, have also been demonstrated for 60 GHz network.

Finally, we have proposed Rate-Distortion-Complexity (R-D-C) control for the upcoming video compression standard High Efficiency Video Coding (HEVC) that allows additional flexibility in terms of complexity. Our method provides a quite accurate control of the complexity and is able to find solutions close to the optimal ones.

In our work we followed some of the main trends in video compression and mobile communications. Some future directions in power-aware mobile multimedia have been pointed out in [140]. One of the challenges in this field is the development of the power management strategies in mobile devices that allow an efficient use of the most critical resource in such systems - battery energy. Another challenge is the R-D-C analysis of video codecs that allows to choose a proper configuration for modern complex encoders like HEVC.

Regarding DVC and its applications, [95] describes the main trends in DVC and concludes that it suits low-complexity applications with low-power consumption at the encoder side, especially when combined with video transmission over noisy channels. With such restrictions DVC represents an appealing solution providing competitive results in terms of rate-distortion performance.

Scalable video coding is gaining its popularity as well. The plan for the development of a scalable extension of HEVC is an additional proof. Due to its properties, scalability looks attractive for video transmission in wireless environments and its possible applications include IPTV [141] and Future Media Internet [142].

The scalability feature poses another issue: it is important to find a method for objective comparison of quality of videos having different temporal and spatial resolutions. Lee et al. [84] confirm that currently most of the subjective metrics are highly dependent on the test data used for metric design. We agree with the authors that it is necessary to perform quality assessment on common validation databases allowing cross-evaluation for benchmarking.

In our opinion one of the most obvious applications that can benefit from low-delay low-power scalable solutions is video surveillance. Most of

the studied approaches, including the quality assessment for video with different frame rates, can be efficiently used in wireless video surveillance systems. Multidimensional quality assessment is almost inevitable in video surveillance that often has to deal with scalable videos. Such systems usually consist of many heterogeneous clients where wireless terminals have time-varying bandwidth capacity. In these conditions, scalable bit streams have advantages against non-scalable solutions [143]. In addition, scalability is useful in video surveillance applications as it allows keeping for long-term storage only some parts of bit stream without degrading visual quality significantly [84].

As new standards and new solutions advance, the potential of wireless video applications grows as well. Many spheres - from education to security, from space science to medicine - are benefiting from the advances in video communications. Many of such applications were unimaginable just few decades ago, but thanks to the technological development we now see and use them in everyday life. It is hard to make predictions about the future research directions, but we think that even if battery capacities of mobile devices will significantly increase in next years, low-complexity approaches will still be desirable as they allow power consumption savings, which gets more and more important in the perspective of limited Earth energy resources.

# Appendices

# Appendix A

# Ph.D. publications

**OBJECTIVE ASSESSMENT OF THE IMPACT OF FRAME RATE ON VIDEO QUALITY**

*Anna Ukhanova, Jari Korhonen,* and *Søren Forchhammer*

Department of Photonics Engineering (DTU Fotonik)
Technical University of Denmark (DTU), Kgs. Lyngby, Denmark

**ABSTRACT**

In this paper, we present a novel objective quality metric that takes the impact of frame rate into account. The proposed metric uses PSNR, frame rate and a content dependent parameter that can easily be obtained from spatial and temporal activity indices. The results have been validated on data from a subjective quality study, where the test subjects have been choosing the preferred path from the lowest quality to the best quality, at each step making a choice in favor of higher frame rate or lower distortion. A comparison with other relevant objective metrics shows that the proposed metric on average provides a more precise correlation with the subjective results.

***Index Terms***— Video quality, frame rate impact, objective metric

**1. INTRODUCTION**

The perceived video quality in today's video applications is a significant part of the Quality of Experience (QoE) for the end users. Therefore, it is very important to measure the possible quality degradations in the system in order to maintain and control the quality of the video data. The methods for quality measurement are divided into two categories: subjective and objective measurements. Subjective quality is often expressed in Mean Opinion Score (MOS), but there are also other possible methods not based on quality scoring, e.g. pairwise comparisons [1]. Whereas subjective quality evaluation provides the most reliable results as quality is estimated by human beings, who represent the end users, it requires more resources and is not suitable for some applications, e.g. in real-time quality monitoring. As the most popular objective metrics, such as peak signal-to-noise ratio (PSNR), do not correlate perfectly with the subjective quality [2], more accurate objective quality metrics are needed for many applications.

As video data also includes the temporal dimension, the quality assessment for video is more demanding than for images. Unfortunately, most of the established video quality metrics do not take frame rate into consideration. However, the impact of frame rate on perceptual video quality has been studied by many researchers. McCarthy, Sasse and Miras studied the effects of quantization vs. frame rate for video sequences with sports content [3]. Their work concluded that on a small screen, high spatial quality is preferred over the frame rate. However, their study did not propose any objective method for quality measurement.

A metric QM based on peak signal-to-noise ratio (PSNR) considering both quantization and frame rate is described in [4]. In case of high frame rates, the metric is dominated by the quantization errors and is close to the PSNR. Otherwise, the PSNR value (measured on the temporally upsampled sequences in case of low frame rate) is compensated depending on the frame rate reduction and the motion speed. However, the compensation is not enough for very low frame rate, as in this case PSNR values for interpolated sequences do not vary significantly for different video quality, which results in almost equal values of the quality metric at low frame rates.

Another quality metric VQMTQ considering both frame rate and quantization artifacts is proposed in [5]. That metric provides a good fit to the measured MOS scores. The model uses sequence-dependent parameters; however, it is possible to predict them based on the characteristics of the video sequences.

Peng and Steinbach proposed a novel full-reference video quality metric STVQM based on PSNR, frame rate, and spatiotemporal activity measures [6]. Their experiments show that they both perform well and the difference between VQMTQ and STVQM metrics is not statistically significant. However, STVQM has some advantages, namely the use of only two standard video activity indicators that can easily be computed, compared to the four parameters with significantly more complex interpretation used in VQMTQ. Moreover, STVQM has no codec-dependent parameters, unlike VQMTQ.

In this paper we propose a novel quality metric for objective quality assessment of the video data taking the impact of the frame rate into account. Our experimental results show that the proposed method allows determining the quality of the video sequences close to the subjective human opinion, and can compete with other recently developed objective quality metrics, such as VQMTQ and STVQM, by offering more constant performance with different contents.

A. Ukhanova, J. Korhonen, S. Forchhammer "Objective assessment of the impact of frame rate on video quality", *International Conference on Image Processing (ICIP)*, Lake Buena Vista (FL), USA, 2012.

## 2. METHODOLOGY

As reference data for this study, we have used the data obtained from the study published in [7]. The purpose of the original experiment was to find the quality optimal path through the plane, where one dimension is the quality measured in terms of peak signal-to-noise ratio (PSNR), and one other dimension is the frame rate. Different combinations of frame rates and PSNRs are represented by nodes, arranged in a form of a two dimensional grid. At each step, test subjects choose between two sequences according to their preference: one with higher frame rate but lower PSNR (the node upwards), and another with lower frame rate but higher PSNR (the node to the right). With this method, it is possible to find the average preferred path from the lowest to the highest frame rate and PSNR, reflecting the perceived relative importance of frame rate and PSNR along the path. The details of the study are omitted due to the lack of space, but interested readers may refer to [7].

Unfortunately, the described method does not produce any subjective quality scores directly. However, we can measure the PSNR difference between two different PSNR levels, and then estimate the perceptual PSNR (PPSNR) difference between two frame rate levels from the relation between test subjects choosing the higher frame rate and those choosing the higher PSNR. The concept is illustrated in Fig. 1. In the illustration, nodes 0, a and b represent video sequences with different combinations of frame rates ($FR_1$, $FR_2$) and PSNR levels ($PSNR_1$, $PSNR_2$). Transition probabilities $p_a$ and $p_b$ denote the proportion of test subjects who prefer to move from node 0 to node a and b, respectively ($p_a + p_b = 1$). PSNR difference between nodes 0 and b is known to be ΔPSNR, and when $p_a$ and $p_b$ are known, we can assume that PPSNR difference ΔPPSNR between nodes 0 and a can be estimated as:

$$\Delta PPSNR = \Delta PSNR \cdot p_a / p_b \qquad (1)$$

The estimate is most reliable when $p_a \sim p_b \sim 0.5$ and becomes less accurate when either $p_a$ or $p_b$ approaches zero. The method can be applied regardless of the sign of ΔPSNR.

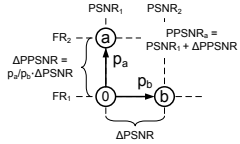



Figure 1. Method for deriving PPSNR values.

The data from [7] contains the preferred paths for five sequences judged by 25 test subjects, across planes with six frame rates and PSNR levels, ie. there are 6x6 nodes in each graph. Each node is assigned with a number of "occupants", ie. test subjects who have traversed through the node in question. For the practical application of the technique described above, we have first initialized the PPSNR values of the nodes with full frame rate with their measured respective PSNR values. Then, we have resolved the PPSNR values cumulatively for the lower frame rates by computing the proportion of "occupants" preferring higher frame rate or PSNR at each step. Naturally, some nodes have too few occupants to give reliable results; these nodes have been discarded from the final results. The most reliable PPSNR values are supposed to be found close to the most popular paths. Figure 2 shows the resulting PPSNR values for sequence "Coastguard".

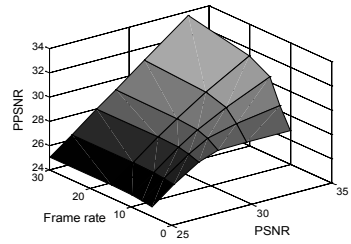



Figure 2. PPSNR values for “Coastguard”.

The next challenge is to find an objective quality index matching with the experimental data obtained in the previous phase as accurately as possible. After attempting with several functions with potential resemblance to the surfaces such as in Fig. 2, we have identified that Eq. (2) gives the most promising results. The parameters of the quality function are *PSNR*, frame rate *FR*, and a content dependent parameter *c*.

$$PPSNR = PSNR \cdot \left(1 - \exp(1 - 10^{8} \cdot FR \cdot PSNR^{-c})\right) \qquad (2)$$

Since parameter *c* is related to the content type, we have used the spatial and temporal activity measures to predict *c*. In this work, we have used slightly modified definitions of the SI and TI indices in [8], denoted as *SA* and *TA*, as defined in [6]:

$$SA = mean_{time}(std_{space}[Sobel(F_n)]) \qquad (3)$$

$$TA = mean_{time}(std_{space}[F_n - F_{n-1}]) \qquad (4)$$

The parameter $c$ is then predicted from $SA$ and $TA$:

$$c = \exp(\alpha + \beta \cdot SA \cdot TA), \quad (5)$$

where $\alpha$ and $\beta$ are model parameters that can be solved by applying linear regression to the experimentally obtained values of ln($c$). The logarithmic space is used instead of linear space, since we got slightly better results using the logarithmic space.

### 3. PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed metric, we have compared our metric with VQMTQ and STVQM, as these are the most relevant comparison points known from the related studies. To make the comparison, we have created the preferred paths as in [7], derived from the quality values produced by the metrics named above. Based on the relative quality differences between adjacent nodes, we have computed the transition probabilities $p_a$ and $p_b$ for each node and then computed cumulatively the relative amount of occupants in each node, starting from the lowest frame rate and quality. The method for computing $p_a$ and $p_b$ is basically the reverse of the method shown in Fig. 1: we assume that $p_a/p_b$ equals to $\Delta VQ_a/\Delta VQ_b$, where $\Delta VQ_a$ and $\Delta VQ_b$ are the differences in quality values between the source node and nodes a and b. Instead of $PPSNR$, $VQ$ can be any quality metric with locally linear behavior. When the objective quality values are known, $p_a$ and $p_b$ can be computed as:

$$p_a = \frac{\Delta VQ_a}{\Delta VQ_a + \Delta VQ_b}, p_b = 1 - p_a \quad (6)$$

As mentioned in Section 2, test subjects make their choice in several steps, starting with the lowest frame rate and PSNR. After the first step, the hypothetical test subjects are divided to nodes (1,2) and (2,1). After the second step, test subjects are distributed among nodes (1,3), (2,2) and (3,1), and so on. If we denote the relative number of occupants in node ($x$,$y$) as $s_{x,y}$, the average indices ($x_i$,$y_i$) after $i$ steps can be computed as follows:

$$\begin{cases} x_i^{obj} = \sum_{x+y=i+2} s_{x,y} \cdot x \\ y_i^{obj} = \sum_{x+y=i+2} s_{x,y} \cdot y \end{cases} \quad (7)$$

From [7], we have the preferred paths based on subjective evaluation for the following sequences in CIF (352x288) resolution: “Akiyo”, “City”, “Coastguard”, “Football”, and “Ice”. The illustrations of the original subjective path for sequences “Coastguard” and “Football” are shown in Fig. 3, together with the paths resolved following the procedure described above, applied to the objective quality indices from VQMTQ, STVQM and the proposed method.

Then, we have compared the subjective path against the objective paths by computing the mean Euclidean distance $D$ between the average subjective position ($x_i^{subj}$, $y_i^{subj}$) and the position ($x_i^{obj}$, $y_i^{obj}$) computed from the objective data at each step $i$:

$$D = mean(\sqrt{(x_i^{obj} - x_i^{subj})^2 + (y_i^{obj} - y_i^{subj})^2}), \quad (8)$$

where position is defined in terms of indices, since FR and PSNR scales are different.

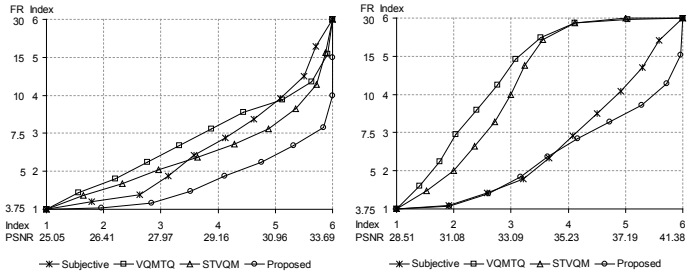
Unfortunately, we only have the subjective paths available for the five abovementioned sequences, and such a small number of data does not allow us to create separate subsets for training and validation. To alleviate this a linear regression was performed for each sequence separately using the “leave-one-out” method, ie. parameters $\alpha$ and $\beta$ required for predicting $c$ were obtained using the known values of $c$ for the other four sequences. The results for the distance comparisons between the subjective preferred path and the objective paths are listed in the Table I for each test sequence. Similar relative results can be obtained by using median instead of mean Euclidean distance.

Table I. Minimum distance comparison for preferred paths obtained by different quality metrics.

| | akiyo | city | coastg. | footb. | ice | **avg.** |
|---|---|---|---|---|---|---|
| VQMTQ | 0.65 | 0.34 | 0.27 | 1.24 | 0.37 | **0.57** |
| STVQM | 0.11 | 0.35 | 0.23 | 1.06 | 0.92 | **0.53** |
| Proposed | 0.11 | 0.69 | 0.54 | 0.20 | 0.49 | **0.41** |

The comparison with the QM metric [4] has not be included, since we have observed that increasing quality and bit rate with low frame rates does not always give increasing QM values, and this is why we have concluded that QM metric cannot predict the quality reliably at low frame rates.

As we can see, on average our proposed metric performs slightly better than VQMTQ and STVQM. However, the performance fluctuates between contents for all the metrics and none of them shows excellent performance on all sequences. For some contents, VQMTQ and STVQM perform better than the proposed metric, but the proposed metric achieves a more constant performance across different contents. The average performance of STVQM and VQMTQ is roughly similar, but VQMTQ has some disadvantages, such as the use of parameters dependent on the codec and content, which are not trivial to compute. We expect that the proposed method can be improved by using subjective data for more sequences, as the use of only four sequences in the “leave-one-out” method for parameter estimation can lead to overemphasis of outliers.




a) "Coastguard" b) "Football"

Figure 3. Preferred subjective and objective paths for sequences "Coastguard" and "Football".

## 4. CONCLUSIONS

Several studies about the impact of frame rate on subjective video quality conclude that the impact is highly dependent on the content. However, most of the established video quality metrics known from the literature do not take frame rate into consideration. In this paper, we have used data from a subjective study where the preferred path from the lowest quality to the best quality is determined by choosing between different frame rate and distortion levels at each hop. We have derived the relative subjective scores from the probability (frequency) of test subjects for choosing each of the two alternatives at each point.

We have proposed a simple metric for assessing the impact of frame rate on quality. In the proposed metric, the PSNR value of the video sequence is multiplied by a factor including frame rate and a content dependent parameter that can be computed from spatial and temporal activity indices SI and TI. We have compared the proposed metric against two other relevant metrics, VQMTQ and STVQM, and observed that in average, the proposed metric is capable of predicting the paths chosen by test subjects more accurately than the other metrics. However, the performance shows a significant dependency on the content. In the future, we plan to improve the proposed metric by involving more extensive subjective experiments, covering a larger variety of contents and higher resolutions.

## 5. ACKNOWLEDGEMENT

The authors thank Dr. Ulrich Reiter from Norwegian University of Science and Technology for providing the subjective data used in this work.

## 6. REFERENCES

[1] J.-S. Lee, F. D. Simone, T. Ebrahimi, "Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding," *IEEE Trans. Multimedia,* vol. 13, no.5, 2011.

[2] B. Girod, "What's Wrong with Mean-squared Error," in Digital Images and Human Vision, A. B. Watson, ed., pp. 207-220, MIT Press, 1993.

[3] J. McCarthy, M. A. Sasse, and D. Miras, "Sharp or Smooth: Comparing the Effects of Quantization vs. Frame Rate for Streamed Video," *Proc. of ACM CHI Hum. Factors Comput. Syst.*, pp. 535–542, Apr. 2004.

[4] R. Feghali, D. Wang, F. Speranza, and A. Vincent, "Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 441–446, Mar. 2007.

[5] Y.-F. Ou, Z. Ma, T. Liu and Y. Wang, "Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, no. 3, 2011.

[6] Y. Peng and E. Steinbach, "A Novel Full-reference Video Quality Metric and Its Application to Wireless Video Transmission," *Proc. of ICIP'11*, Sep. 2011.

[7] J. Korhonen, U. Reiter, and J. You, "Subjective Comparison of Temporal and Quality Scalability," *Proc. of QoMEX'11*, Sep. 2011.

[8] ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," International Telecommunication Union, Sep. 1999.

Computer Communications 35 (2012) 1695–1706

Contents lists available at SciVerse ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

# Power consumption analysis of constant bit rate video transmission over 3G networks

Anna Ukhanova [a,*], Evgeny Belyaev [b], Le Wang [c], Søren Forchhammer [a]

[a] DTU Fotonik, Technical University of Denmark, Ørsteds Plads, B. 343, 2800 Kgs. Lyngby, Denmark
[b] Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
[c] Aalto University, P.O. Box 13000, FI-00076 Aalto, Finland

ARTICLE INFO

*Article history:*
Available online 27 May 2012

*Keywords:*
3G
Video transmission
Constant bit rate
Power consumption modeling
Power saving

ABSTRACT

This paper presents an analysis of the power consumption of video data transmission with constant bit rate over 3G mobile wireless networks. The work includes the description of the radio resource control transition state machine in 3G networks, followed by a detailed power consumption analysis and measurements of the radio link power consumption. Based on this description and analysis, we propose our power consumption model. The power model was evaluated on a smartphone Nokia N900, which follows 3GPP Release 5 and 6 supporting HSDPA/HSUPA data bearers. We also propose a method for parameter selection for the 3GPP transition state machine that allows to decrease power consumption on a mobile device taking signaling traffic, buffer size and latency restrictions into account. Furthermore, we discuss the gain in power consumption vs. PSNR for transmitted video and show the possibility of performing power consumption management based on the requirements for the video quality.



## 1. Introduction

Wireless networks have gone through an extensive development throughout the recent years. After the Internet, in the 1990s, came the second generation wireless systems with additional features for communication by means of cellular phones. With the move from analog (1G) to digital transmission (2G), the capacity of the cellular systems have been significantly improved. Further demand for greater bandwidth resulted in the development of third generation mobile telecommunications (3G), providing a mobile broadband access to handheld devices.

The number of mobile cellular subscriptions has also been rapidly growing in the last decade, and now the rough estimate has almost reached 6 billions. Nowadays 3G systems are also becoming ubiquitous: the number of active mobile-broadband subscriptions reaches almost 1.2 billion [1]. As wireless communication technology advances, in new generation mobile phones 3G networks provide a large amount of various services including high data rate Internet access, videoconferencing, global positioning, high quality music and video downloading, and gaming capabilities.

Mobile TV and video-on-demand services are also rising in popularity and expected to be a significant driver for the wireless consumer industry. Consumers demand for video data, quality of experience of mobile multimedia and its usability is pushing the development of Mobile TV and related services. Therefore, video transmission for mobile terminals is a major application in the 3G and beyond systems and may play a key role in their success [2]. According to Cisco research, "two-thirds of the world's mobile data traffic will be video by 2015" [3]. As a consequence, these additional features are bringing the power consumption of mobile phones to the level of desktop computers. However, the lack of a constant power supply and limited battery capacity [4] pose strict limits to the overall power consumption of the device. Therefore, minimization of the power consumption of wireless devices is a great challenge for the entire electronic industry, at all system levels. Hence, an intense research in this field has focused on power management [5–7].

In our work we analyze the power consumption in case of uplink transmission, show how the power consumption depends on the transmission parameters and explain how it can be reduced by optimizing the power management policy. We analyze the cases with buffer and latency restrictions and show how to choose transmission parameters in these scenarios. Due to the growing interest in various mobile video applications, the analysis is then extended to the case of constant bit rate video transmission. Experimental results show that in this case the proposed solution allows to save power on video transmission if compared to the conventional approach.

Our paper starts with a review of some prior work on the power consumption in wireless networks in Section 2. Then we present

* Corresponding author. Tel.: +45 45256567, mobile: +45 50679631; fax: +45 45936581.
*E-mail address:* annuk@fotonik.dtu.dk (A. Ukhanova).

0140-3664/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comcom.2012.05.010

A. Ukhanova, E. Belyaev, L. Wang, S. Forchhammer "Power consumption analysis of constant bit rate video transmission over 3G networks", *Computer Communications Journal*, Vol. 35, Iss. 14, 2012.

the 3G mobile transmission system in Section 3, where we explain the idea of state machine applied to mobile devices and describe the communication between transmitter and receiver. In Section 4 the state power model is provided, with particular focus on Cell_FACH (Cell Forward Access Channel) and Cell_DCH (Cell Dedicated Channel) states. We propose a model of power consumption in Cell_DCH state and show that power consumption is related to the packet size and transmission interval. In Section 5 we propose a method of parameter selection for the 3GPP transition state machine that allows to decrease power consumption on the mobile device in case of constant bit rate data transmission. Section 6 presents the power consumption for constant bit rate video transmission and provides the results for trade-off between video quality and power consumption for several standard test video sequences.

## 2. Power consumption in wireless networks

There exists a large number of approaches for power management in communication networks for handheld devices and they have been studied throughout the recent years.

An overview of the methods to reduce the large and growing energy consumption of the Internet was provided in [8]. However, since that time the Internet has been adopted and widely used in many other devices other than networked desktop computers, and deployment of next generation networks in the mobile devices has significantly contributed to it.

Yeh et al. [9] as well as Sklavos and Touliou [10] provided an analysis of power consumption in 3G networks. Markov chains are used in [9] to analyze the 3GPP transition state machine and examine how different timer values affect the power consumption of the device. The analysis is based on several assumptions about the device power consumption in different states. Sklavos and Touliou [10] considered different units of the mobile phone (such as memory, display, cellular engine), analyzed power consumption on each of them and addressed different power management techniques for different units, thereby focusing on the device power consumption. They also state the need for power control techniques, e.g. based on method of idling the modules that are not used. A similar approach of dividing the processes in stages and applying different power saving mechanisms to them is described in [11]. For the reception stage, which requires most of the power, the solution is to send multimedia data in larger amounts at less frequent intervals by hiding the traffic temporarily from a mobile station, thus allowing it to sleep longer and achieve power savings. Korhonen and Wang [12] also proved that in IEEE 802.11 power of the receiver can be saved if data packets are transmitted as bursts.

Perrucci [5] provided a broad overview of existing approaches and studied the strategies for efficient use of the wireless communication in three main focus areas, namely Cross Layer, Overlay Networks and Cooperation. The energy savings on mobile devices are achieved by choosing the most efficient available network, using wake-up systems (described in detail in [13]) and sharing cellular links between mobile users. Another analysis of approaches for achieving energy-efficient web access on mobile devices is provided in [14].

Two good overviews of the past and future research directions in the field of power management for mobile networks are given in [6,7]. In [6] the mobile device and battery capacity evolutions, multiaccess nature of modern mobile devices and the respective implications for power management are described, as well as a proposal of an information-centric approach to networking, which allows audio/video streaming to be transformed from an energy-heavy network service to a lightweight one. However, the last would require an adoption of a new networking paradigm and a change in the way the information is distributed to its intended recipients. Zhang et al. [7] provided a survey of many issues related to power-aware mobile multimedia, such as power-management for mobile devices, rate-distortion-complexity optimized video codec design, and computational complexity and power aware cross-layer design and optimization. The authors have pointed out the challenges and the corresponding future research directions in power-aware video coding and power-aware video delivery, such as power management in mobile devices, rate-distortion-complexity analysis of video codecs and network information feedback and cross-layer signaling.

Considering wireless video transmission from mobile devices, it is necessary to take into account that some energy has to be spent on the video capturing and compression. On average, a mobile camera may consume less than 60 mW for CIF resolution video at 30 Hz [15]. According to [16], the power needed for compression is also lower than the one needed for the transmission. Nevertheless, cross-layer approaches could be applied here as well and work like [17] discusses joint power control and bit allocation for video transmission in wireless networks. A good overview of works on this topic is also given in [18].

Chen et al. [19] described the fundamental trade-offs in wireless networks, such as deployment efficiency – energy efficiency, spectrum efficiency – energy efficiency, bandwidth – power, and delay – power trade-offs. Kim et al. [20] and Vuyst et al. [21] explored the trade-offs between delay and power consumption for sleep-mode operation in mobile WiMAX. In [22,23] the possibility for trade-off between video quality and power saving in the receiver was demonstrated. In [22] the mobile broadcast standard DVB-H along with JPEG2000 for video encoding was used to show how to allow receivers to control the level of power consumption depending on the priorities. As any other scalable codec could be used instead of JPEG2000, the work was further extended to the Scalable Extension of H.264/AVC [23].

Unlike the work in [22,23], this time we consider the case of power consumption for the uplink transmission only and thereby focus on the device. However, the main idea – that it is possible to control the trade-off between the level of power consumption and video quality – has been kept. Moreover, we also take the delay – power trade-off into account and show how it influences the achieved power savings. We provide a power model for devices for 3G systems, and perform power-aware cross-layer optimization by controlling the transmission parameters and adjusting the time periods spent by the mobile device in the state with active data transmission. We also show the potential for power management for video transmission that is content-dependent as different videos can have different rate-distortion characteristics.

## 3. Mobile transmission system overview

3G systems provide global communication with various services including telephony, messaging and access to Internet. 3G networks consist of three domains: Core Network (CN), UMTS Terrestrial Radio Access Network (UTRAN) and User Equipment (UE). UE interoperates with Base Station (called Node B). The Radio Resource Control (RRC) handles the control plane signaling between the UEs and the UTRAN. For efficient use of radio resources and power consumption control, RRC introduces a state machine for UE [24].

### 3.1. State machine

There are five states in the RRC: Idle, Cell_FACH (Cell Forward Access Channel), Cell_PCH (Cell Paging Channel), Cell_DCH (Cell Dedicated Channel) and URA_PCH (Utran Registration Area Paging Channel). Cell_PCH and URA_PCH can be considered as low power

states, which consume only around 30 mW. The state of Cell_FACH consumes around 400 mW and the state of Cell_DCH consumes around 800 mW, according to our measurements on a smartphone Nokia N900. URA_PCH is very similar to Cell_PCH, although some vendors have not implemented it in their solution. In our work we consider these two states largely equivalent.

The power consumption in Cell_FACH is roughly 50% of that in Cell_DCH, and Cell_PCH state uses about 1–2% of the power consumption of Cell_DCH state [25]. Each state is now described in more detail.

**Idle**. In this mode UE does not communicate with the network although it does listen for broadcast messages. So it does not have a RRC Connection, but UE can still have an IP address and be reached by paging. In this state the mobile device consumes the least amount of power.

**Cell_PCH**. In this state the channel is shared by all mobile devices so the inclusion of an additional mobile device does not have any impact on the network. UE monitors paging messages from the Radio Network Controller (RNC). As in the Idle state, the power consumption is very small. In this state no dedicated physical channel is allocated to the UE, so no uplink activity is possible.

**Cell_FACH**. In the Cell_FACH the mobile device communicates with the network via a shared channel. A few bits of data can be transmitted at a relatively low data rate, on the order of up to 16 kbps in the uplink. The maximum amount of transmission data also depends on the overall loading of the common channels. At the same time the UE continuously monitors a FACH in the downlink. The mobile device power consumption is higher than it is in Idle or Cell_PCH states.

**Cell_DCH**. The mobile device is allocated a dedicated transport channel both in downlink and uplink. It consumes the most network resources and the impact on the battery is at the very high level.

### 3.2. Communication between transmitter and receiver

It is UE that always initiates the RRC connection, then the establishment and the release are handled by the RRC protocol. UE starts working in Idle state, when an RRC connection has not yet been established. Only one RRC connection is used at any time between the UE and the network. When an RRC connection has been established between UE and Node B together with RNC, the Idle state switches to the RRC Connected mode.

To be more precise, from Idle mode through establishment of an RRC connection the UE enters the Cell_DCH state. Further it can be moved by explicit signaling from Cell_DCH to other states. The UE does not generally listen to the broadcast channel in this state. If Node B allocates to UE a common or shared traffic channel (i.e., the channel is shared by several UEs), it enters Cell_FACH state. The data communication activities can only be performed in these two states.

Depending on the activities of the UE and traffic volume, states could be changed. Signaling messages (radio bearer configuration messages) are sent between UE and Node B when states are changing. Three timers are used to detect when a mobile device should move to a lower power state in case of inactivity. These inactivity timers T1, T2, and T3 are managed by RNC.

T1 is used in Cell_PCH. After T1 seconds (usually a very long timer), the RRC connection will be released and the state will be changed to Idle. T2 is an inactivity timer determining how long the 3G device should remain in Cell_FACH state without any activity. Timer T3 is used within Cell_DCH state and refers to the inactivity period after which the 3G device enters Cell_FACH.

As for the signaling traffic i.e. messages between UE and Node B, it is necessary to note that 3G was designed and implemented to support large amounts of data traffic (like long, uninterrupted data sessions, video conferencing, etc.). But according to Thelander, CEO and founder of Signals Research Group (SRG), the reality is that "signaling traffic is outpacing actual mobile data traffic by 30–50%, if not higher" [26].

## 4. Power consumption analysis of transition states

In this section, the power consumption of each state of the RRC is analyzed. We focus on modeling the power consumption in Cell_DCH state. Then, our proposed model is compared with real measurements and a reference model which thoroughly analyzes power consumption for UDP traffic in 3G network and provides practical power model for the consumption.

### 4.1. The influence of packet sending intervals and packet size on power consumption of Cell_DCH state

Each RRC state requires varying power to maintain operation and differs when generating signals transferred between UE and RNC to establish, maintain and release connections as well as transmit/receive data across the air interface. Compared to Cell_FACH state, which is only applicable for transferring relatively small quantities of data, Cell_DCH state gives potential for UE to transfer large quantities of data and, thus, is the state where most of the data communication happens, as specifically described in the previous section. It is the most interesting state to examine the effect of packet sending intervals and packet size on power consumption of UE radio interface.

Generally speaking, the size of a transport block specifies the maximum payload that can be transmitted within each Transmission Time Interval (TTI), which decides the maximum packet sending or receiving rate. These two parameters together influence the maximum throughput and packet sending or receiving pattern in the Physical layer. From the Transport layer perspective, application's traffic pattern – namely packet sending or receiving interval and packet size – directly decides underlaying layer's behavior regarding the size of transport block set and transmitting interval. Specifically, if the number of bits in a TTI is larger than the maximum size which one physical block can contain, segmentation is performed and the over-sized bits are sent during the next TTI. If the packets are generated with an interval which is less than TTI, the transmitting interval is decided by TTI. Otherwise, it is decided by the packet generating interval of the application. The power consumption of a radio interface increases proportionally to the number of transport block sets sent and received over one radio interface.

Fig. 1 demonstrates the power consumption of packet transmission on a smartphone Nokia N900 in downlink. A traffic generator is used to generate UDP packets in different intervals as shown in the figure. Each peak corresponds to the power consumption of transmitting one transport block set. As it can be seen, the number of peaks directly influences the power consumption of the radio interface. A faster receiving interval leads to a larger number of peaks, and thus higher power consumption. In our experiments, we also observed the same phenomena in uplink on Nokia N900. Among the peaks, wide ones are due to the power consumption of daemon processes of 3G modem and excluded from calculation in the following subsections. The power consumption was measured in a stable environment. Otherwise, when the signal quality of UE drops off or the UE moves away from the Node B, power control mechanisms in 3G network have to increase transmission power in order to keep the received uplink Signal-to-Interference Ratio (SIR) at a given SIR target. Moreover, the degradation of signal quality also increases MAC layer retransmission, which leads to more energy consumption for successful transmission of one bit.

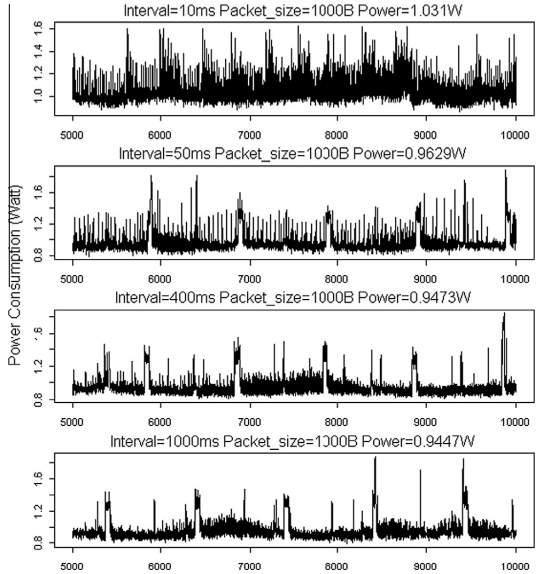



**Fig. 1.** Power consumption of Nokia N900 in downlink.

The measurements were done three times. The average variance of the transmission power consumption (sending and receiving) is $5.7 \cdot 10^{-6}$.

### 4.2. Power consumption modeling of Cell_DCH state

The power consumption of sending or receiving packets in Cell_DCH state mainly consists of three sources: the power consumption of Cell_DCH state defined as $P_{DCH}$ in watt, the power consumption of sending or receiving packets which is defined as $P_{peak}$ in watt demonstrated as power peaks in Fig. 1, and the power consumption of encapsulating or decapsulating packets which is defined as a function of packet size $P_{enc}(s)$, where $s$ is the size of packet in bytes. Therefore, the power consumption of UE to send or receive packets in state Cell_DCH can be reformulated as follows:

$$P = P_{DCH} + P_{peak} + P_{enc}(s). \quad (1)$$

UE requires resource and power consumption to maintain the state. Thus, $P_{DCH}$ is the minimal power consumption for the UE to stay in state Cell_DCH, which includes the power consumption of reception of control signals and is considered as an approximately fixed value since most of the traffic is data traffic.

Besides, we define $I$ as the UDP packet sending interval in Transport layer in ms, which is the interval of sending packets from UE to Node B or the interval of sending packets from Node B to UE.

The power consumption for encapsulation or decapsulation $P_{enc}(s)$ for packet size $s$ is assumed to be linear and proportional to the size of the packet since more computational power is required to process bigger packets yielding higher power consumption. Since $P_{DCH}$, $P_{peak}$ and total power consumption $P$ are measurable, the power component $P_{enc}(s)$ can be calculated and then used to derive the linear model.

The number of transport blocks needed for sending one IP packet is determined by the size of the packet and maximum number of bits for a transport block per TTI defined in different 3GPP releases. Here we define the number of transport blocks as

$$N = \left\lceil \frac{s}{MTBS} \right\rceil, \quad (2)$$

where MTBS is Maximum Transport Block Size.

Let $E_{peak}$ denote the energy consumption of sending or receiving one peak in Joule. As analyzed above, the power consumption of peaks changes more or less linearly with the number of the transport blocks. When more than one transport block is needed to send or receive one IP packet, the time spent on processing this packet is $N \cdot \tau$, where $\tau$ is the value of TTI. Normally, packet sending interval $I$ is much larger than the packet processing time. Thus,

$$P_{peak} = \frac{N}{I} \cdot E_{peak}, \quad \text{when} \quad I \gg N \cdot \tau. \quad (3)$$

Then combining (1)–(3), the power consumption in Cell_DCH state can be written as

$$P = P_{DCH} + \frac{E_{peak}}{I}\left(\left\lceil \frac{s}{MTBS} \right\rceil\right) + P_{enc}(s). \quad (4)$$

Eq. (4) formulates the power consumption of one connection of uplink or downlink traffic. This model shows that power consumption can be determined by the number of peaks, namely set by packet sending or receiving interval $I$ and packet size $s$, which is the influential factor of power consumption of the radio interface. This model can be extended to formulate the power consumption for multiple connections by counting the amount of bits sent or received in a certain interval. As long as mobile devices are capable to record transmitted and received packet intervals and sizes, the proposed model can be extended to estimate power consumption of the radio interface in runtime.

### 4.3. Experimental setup

In order to have sufficient measurement accuracy and an uninterrupted power source, the battery of the N900 was replaced with a battery adapter, which was serially connected to a 4.1 V DC power supply and a 0.1 Ohm resistor. A NI cRIO-9215 was then used as a data logger to record voltage fluctuations of the N900 at a sample rate of 1000 sample/s. A Linux traffic generator was also used to generate packets with various packet sizes and sending intervals. UDP traffic was generated instead of TCP to avoid TCP hand-shake and retransmissions. The packet sending interval ranged from 10 ms to 1000 ms to avoid UE switching to Cell_FACH state. The packet size ranged from 10 bytes to 1500 bytes, which is the typical Maximum Transmission Unit (MTU) for Ethernet. When measuring the power consumption of sending packets, the UDP packets generated on the N900 were sent to Node B via uplink, and then forwarded to the Linux server. When measuring the power consumption of receiving packets, the UDP packets were generated on the Linux server and received by the N900 in downlink.

### 4.4. Evaluation of power models

The power model was evaluated on Nokia N900, which follows a 3GPP Release 5 and 6 supporting HSDPA/HSUPA data bearers. The measured average energy consumption of peak $E_{peak}$, average power consumption in Cell_DCH $P_{DCH}$ and related parameters needed for (1) are listed in Table 1. Through (4), the power consumption of sending or receiving packets is thereby decided by $I$ and $s$.

In order to evaluate our model, we use another power consumption model, which is based on the assumption that power consumption is linear in the data rate $r$ [27]. The referenced model designates two independent variables.

- Time of data communication.
- Amount of data sent or received during the communication.

**Table 1**
Parameters for the experiment applied on Nokia N900.

| | Uplink | Downlink |
|---|---|---|
| UE category | HSUPA category 5 | HSDPA category 5 |
| TTI | 10 ms | 2 ms |
| Maximum Transport Block Size (MTBS) | 20000 bits | 7298 bits |
| Data rate | 2 Mbit/s[a] | 2 Mbit/s |
| $E_{peak}$ | 0.4532E−3 J | 0.4435E−3 J |
| $P_{DCH}$ | 0.8556 W | 0.8478 W |

[a] *Note*: The uplink data rate is limited to 2 Mbit/s due to the type of data packet. The maximum data rate is 3.65 Mbit/s.

The energy consumption of UDP-type session, therefore, is expressed as follows.

$$E = t \cdot (r_t + r \cdot r_d) + c, \quad (5)$$

where $r_d$ is the energy consumption rate for data in Joule/KByte, $r_t$ is power consumption in Watt, $r$ is data rate, $t$ is transmission time in second and constant $c$ represents the offset term of the energy consumption, which is independent of the process duration. The model is introduced as a general equation of energy consumption to describe UDP-type of data session. In order to compare it with our power consumption model, the equation needs to be divided by time $t$. The power consumption model can be formulated as a linear equation.

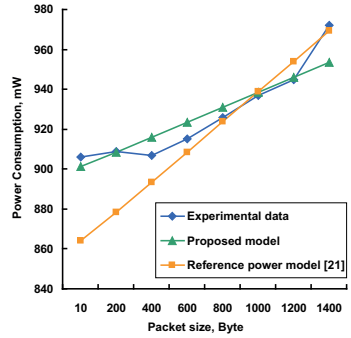



**Fig. 2.** Uplink power consumption for experiment, proposed model and reference model for time interval = 10 ms.

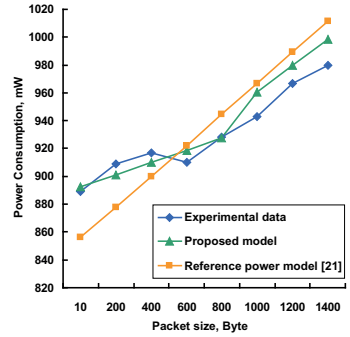



**Fig. 3.** Downlink power consumption for experiment, proposed model and reference model for time interval = 10 ms.

$$P = r_t + r \cdot r_d. \quad (6)$$

The fitted values of the proposed model and reference model, and the measured values from experiments are shown in Figs. 2 and 3. To compare two models, rate $r$ in the reference model is expressed as the product of packet size $s$ and packet sending frequency $I^{-1}$ such that two models share the same coordinate. Our proposed model shows a better approximation than the reference model. We also compared the two models with their Mean Absolute Percentage Error (MAPE) for downlink and uplink power consumption, which is defined as follows.

$$MAPE = \frac{1}{n} \cdot \sum_{i=0}^{n} \frac{|A_i - F_i|}{A_i}, \quad (7)$$

where $n$ is the number of measured values of power consumption, $A_i$ is the actual value and $F_i$ is the fitted value. The MAPEs of the proposed model is 5.0% for downlink and 4.7% for uplink power consumption, which is about 28% more accurate than that of the reference model.

## 5. Uplink power consumption analysis of transition state machine

Section 5 provides the uplink power consumption analysis of the transition state machine (TSM) described in Section 3. As we consider constant bit rate sources (in contrast to http-like data sources, modeled e.g. as Poisson source traffic [28]), it is easy to predict the source rate and find optimal parameters for TSM to minimize the uplink power consumption. This section starts with an evaluation of power consumption for ideal transition state machine. Further this ideal TSM is used for parameter selection for 3GPP TSM, and the differences between these two TSMs are also described. The paper continues with the analysis of the TSM and parameter selection in case for different limitations: the limitation for the amount of signaling traffic (Section 5.3), the limitations for the amount of signaling traffic and buffer size (Section 5.4), and the limitations for the amount of signaling traffic and buffering latency (Section 5.5). The corresponding results for power savings in these cases are presented on the figures in this Section.

### 5.1. Power consumption for ideal transition state machine

Let $P$ denote the uplink power consumption of mobile device, $r$ the data source bit rate and $c$ the channel bit rate. The optimal TSM has to correspond to the following optimization task:

$$\begin{cases} \text{minimize } P, \\ c \geqslant r. \end{cases} \quad (8)$$

Let us define $B_{RLC}$ as the uplink data buffer of UE, $B^T_{RLC}$ as the buffer threshold and $t_{inact}$ as the inactivity time. Then, the following ideal TSM can be introduced:

**State 1** (Cell_PCH).
If activity detection, then go to State 2,
else go to State 1.
**State 2** (Cell_FACH).
If $B_{RLC} > B^T_{RLC}$, then go to State 3,
else if $t_{inact} > T_2$, then go to State 1,
else go to State 2.
**State 3** (Cell_DCH).
If $B_{RLC} = 0$, then go to State 2,
else go to State 3.

Let us define $p_1, p_2$ and $p_3$ as the power consumption in State 1, State 2 and State 3 respectively and $c_2$ and $c_3$ as channel rate in State 2 and State 3. The power consumption of this TSM depending on the data source bit rate $r$ can be described as follows. If data rate $r = 0$, then UE is always working in State 1 and power consumption is $p_1$. If $0 < r \leqslant c_2$, then UE is always working in State 2 and has power consumption $p_2$. If $r \geqslant c_3$, then UE is always working in State 3 and has power consumption $p_3$. If $c_2 < r < c_3$ then there is buffer accumulation in State 2 and buffer emptying in State 3. Accumulation time in State 2 is

$$t_2 = \frac{B^T_{RLC}}{r - c_2}. \quad (9)$$

Emptying time in State 3 is

$$t_3 = \frac{B^T_{RLC}}{c_3 - r}. \quad (10)$$

Finally, power consumption for the ideal TSM is:

$$P = \begin{cases} p_1, & \text{if } r = 0, \\ p_2, & \text{if } 0 < r \leqslant c_2, \\ \frac{t_2}{t_2+t_3} \cdot p_2 + \frac{t_3}{t_2+t_3} \cdot p_3, & \text{if } c_2 < r < c_3, \\ p_3, & \text{otherwise}. \end{cases} \quad (11)$$

**Theorem 1.** *The state machine described above is a solution of the optimization task (8) for data source bit rate $r \in (c_2, c_3)$.*

**Proof.** From (11) it follows that for ideal TSM the channel rate $c = r$ for $r \in (c_2, c_3)$. Let us assume that another TSM with power consumption $P' < P$ exists. It is possible only if accumulation time $t'_2$ in State 2 for this TSM is more than the accumulation time $t_2$ in State 2 for ideal TSM. But in this case channel rate $c'$ for this TSM will be less than $r$. It means that this TSM does not exist. □

### 5.2. Power consumption for the 3GPP transition state machine

The ideal TSM can be used to select the parameters for 3GPP TSM [29]: $B^T_{RLC}$ which was defined in previous subsection and timer $T_3$. Then 3GPP TSM differs from the ideal TSM only in State 3, that can be described as follows (see Fig. 4):

**State 3** (Cell_DCH).
If $r < c_3$ more than $T_3$ sec, then go to State 2,
else go to State 3.

If data rate $r = 0$, then UE is always working in State 1 and power consumption is $p_1$. If $0 < r \leqslant c_2$, then UE is always working in State 2 and has power consumption $p_2$. If $r \geqslant c_3$, then UE is always working in State 3 and has power consumption $p_3$. If $c_2 < r < c_3$ then there is buffer accumulation in State 2 and buffer emptying in State 3. Finally, a power consumption model for 3GPP TSM is given by (11) with accumulation time in State 2

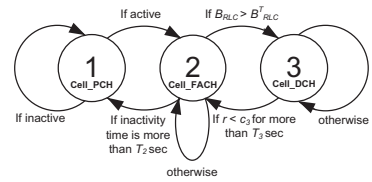



**Fig. 4.** 3GPP transition state machine.

$$t_2 = \frac{\min\{B^T_{RLC}, (c_3 - r)\cdot T_3\}}{r - c_2} \quad (12)$$

and emptying time in State 3 $t_3 = T_3$.

From Theorem 1 follows that if parameters $T_3$ and $B^T_{RLC}$ of 3GPP TSM satisfy the equation

$$B^T_{RLC} = (c_3 - r)\cdot T_3, \quad (13)$$

then 3GPP TSM is identical to ideal TSM and it is a solution of the optimization task (8) too.

Figs. 5 and 6 show power consumption for TSM with experimentally obtained parameters and parameters for carrier 1 listed in [24] and compared with TSM with proposed parameters selected using (13) for constant bit rate data. In fact, the parameters depend on the operator; some additional data and discussions could be found in [30]). These figures illustrate that with the proposed parameter selection the power consumption is growing linearly with the growth of bit rate, which allows a significant decrease of uplink power consumption of UE.

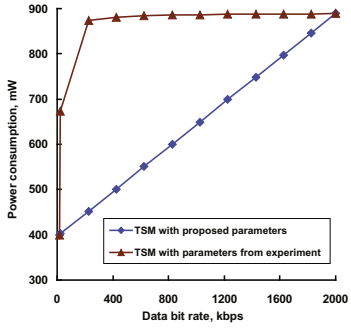



**Fig. 5.** Uplink power consumption for transition state machine for constant bit rate data transmission, $c_2 = 16$ kbps, $c_3 = 2000$ kbps, $B_{RLC} = 8$ kB, $T3 = 9$ s, $p_2 = 400$ mW, $p_3 = 890$ mW.

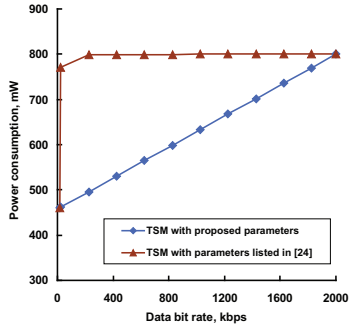



**Fig. 6.** Uplink power consumption for transition state machine for constant bit rate data transmission, $c_2 = 16$ kbps, $c_3 = 2000$ kbps, $B_{RLC} = 543$ B, $T3 = 5$ s, $p_2 = 460$ mW, $p_3 = 800$ mW.

### 5.3. Parameter selection taking signaling traffic into account

As was mentioned in Section 3, one of the problems for 3G networks is the increasing amount of the signaling traffic [31,32,24], as each time UE moves between the various RRC states it exchanges signaling messages with the mobile network for mobility and radio resource management. In this subsection we show the possibility of controlling the total signaling traffic on the base station.

Let us define $n_{23}$ as the number of signaling messages needed to be transmitted from UE to base station for the transition from State 2 to State 3 and $n_{32}$ as the one needed for the transition from State 3 to State 2 and $n_s$ as the number of signaling messages per second for UE. Then, optimization task (8) can be modified as follows:

$$\begin{cases} \text{minimize } P, \\ c \geqslant r, \\ n_s \leqslant N^*_s, \end{cases} \quad (14)$$

where $N^*_s > 0$ is a maximum allowed number of signaling messages per second for UE.

If $c_2 < r < c_3$ the UE works in State 2 and State 3 only. In case of constant bit rate transmission all working time can be divided into the equal intervals $t_2 + t_3$ seconds (see Fig. 7). The number of transmitted signaling messages in each interval is a sum of transition signaling messages for both states $n_{23} + n_{32}$. Therefore the number of signaling messages per second can be calculated as

$$n_s = \frac{n_{23} + n_{32}}{t_2 + t_3} \leqslant N^*_s. \quad (15)$$

Combining (9), (10) and (15), it follows:

$$B^T_{RLC} \geqslant \frac{(n_{23} + n_{32})\cdot(r - c_2)\cdot(c_3 - r)}{N^*_s \cdot (c_3 - c_2)}. \quad (16)$$

The maximum needed $B^T_{RLC}$ is

$$\max_r B^T_{RLC} = \frac{(n_{23} + n_{32})\cdot(c_3 - c_2)}{4\cdot N^*_s}. \quad (17)$$

From (17) follows that 3GPP TSM is a solution of (14) if

$$\begin{cases} B^T_{RLC} = \dfrac{(n_{23} + n_{32})\cdot(c_3 - c_2)}{4\cdot N^*_s}, \\ T_3 = \dfrac{B^T_{RLC}}{c_3 - r}. \end{cases} \quad (18)$$

In case of constant bit rate data transmission, (18) allows to control the total network signaling traffic on the base station by selecting $N^*_s$ for each UE.

### 5.4. Parameter selection taking signaling traffic and UE buffer size restrictions into account

In real applications the UE buffer size can be limited depending on device properties. In this case, it is important to comply with this limitation while choosing the parameters for the transmission in the proposed method. Taking into account buffer size restrictions, the optimization task (14) can be modified as follows:

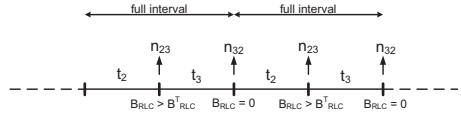



**Fig. 7.** Amount of signaling traffic for a transmission interval.

$$\begin{cases} \text{minimize } P, \\ c \geqslant r, \\ n_s \leqslant N_s^*, \\ B_{RLC}^T \leqslant B_{RLC}^{max}, \end{cases} \tag{19}$$

where $B_{RLC}^{max}$ is the maximum possible UE buffer size.

Taking into account buffer size restrictions, the working time of UE in State 2 is

$$t_2 = \frac{B_{RLC}^{max}}{r - c_2}. \tag{20}$$

At the same time the working time of UE in State 3 can be written as

$$t_3 = \frac{B_{RLC}^{max}}{c_3 - r} + \Delta t_3, \tag{21}$$

where $\Delta t_3 \geqslant 0$ is an additional time needed to provide the maximum allowed number of signaling messages per second $N_s^*$.

Taking into account (15), (20) and (21), the additional time has to satisfy the following inequality:

$$\Delta t_3 \geqslant \frac{n_{32} + n_{23}}{N_s^*} - \frac{B_{RLC}^{max}}{r - c_2} - \frac{B_{RLC}^{max}}{c_3 - r}. \tag{22}$$

To minimize the power consumption and provide $c = r$ the inequality in (22) should be written as equality. As a result, for $c_2 < r < c_3$ 3GPP TSM will be a solution of (19) if

$$\begin{cases} B_{RLC}^T = B_{RLC}^{max}, \\ T_3 = \frac{B_{RLC}^{max}}{c_3 - r} + \max\left\{0, \frac{n_{32}+n_{23}}{N_s^*} - \frac{B_{RLC}^{max}}{r - c_2} - \frac{B_{RLC}^{max}}{c_3 - r}\right\}. \end{cases} \tag{23}$$

Figs. 8 and 9 show the comparison between the power consumption for systems with TSM without any restrictions and with different buffer limitations. Formula (23) shows that with the increase of restricted $N_s^*$ the time portion of working in State 2 becomes closer to the maximum possible value without restrictions (18). For low values of $N_s^*$ there is an opposite situation, and the UE should stay longer in State 3, therefore the gain in power consumption savings is lower. For $N_s^* = 2$ only the curve for $B_{RLC}^{max} = 200$ kB differs from the one without restrictions. That means the value of $N_s^*$ is high enough to adapt the parameter values, unless the buffer becomes too small to keep the UE in State 2 long enough. Thus, the result of the gain in power consumption savings with buffer restrictions depends on the current situation in the network and limitations on $N_s^*$, imposed by the network operator.

### 5.5. Parameter selection taking signaling traffic and buffering latency restrictions into account

Another important restriction is buffering latency $\Delta T$, which has a significant meaning for real-time wireless communications. Necessary requirement for conversational services is an end-to-end delay smaller than 250 ms [33], but, according to [34], the desired delay in real time video applications is approximately 40 ms. For non-real time video applications the limit value is higher and should not be greater than 10 s, as stated in [35]. This limitation

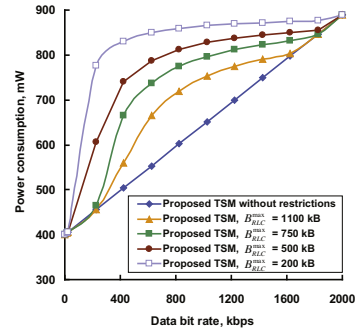



**Fig. 8.** Power consumption with buffer restrictions and $N_s^* = 0.4$.

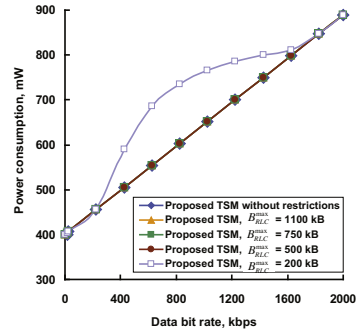



**Fig. 9.** Power consumption with buffer restrictions and $N_s^* = 2$.

for buffering latency also influences our method for efficient power savings.

Taking buffering latency restriction into account, the optimization task (14) can be modified as follows:

$$\begin{cases} \text{minimize } P, \\ c \geqslant r, \\ n_s \leqslant N_s^*, \\ \Delta T \leqslant L, \end{cases} \tag{24}$$

where $L$ is a buffering latency restriction.

Taking this restriction into account, the working time of UE in State 2 is

$$t_2 = L. \tag{25}$$

At the same time the working time of UE in State 3 can be written as

$$t_3 = \frac{L \cdot (r - c_2)}{c_3 - r} + \Delta t_3, \tag{26}$$

where $\Delta t_3 \geqslant 0$ is an additional time needed to provide the maximum allowed number of signaling messages per second $N_s^*$. Taking (15) into account, $\Delta t_3$ has to satisfy the following inequality:

$$\Delta t_3 \geqslant \frac{n_{32} + n_{23}}{N_s^*} - L - \frac{L \cdot (r - c_2)}{c_3 - r}. \tag{27}$$

In order to minimize the power consumption and provide $c = r$, the inequality in (27) should be written as equality. As a result, for $c_2 < r < c_3$ 3GPP TSM will be a solution of (24) if

$$\begin{cases} B_{RLC}^T = L \cdot (r - c_2), \\ T_3 = \frac{B_{RLC}^T}{c_3 - r} + \max\left\{0, \frac{n_{32}+n_{23}}{N_s^*} - L - \frac{L \cdot (r - c_2)}{c_3 - r}\right\}. \end{cases} \tag{28}$$

Figs. 10 and 11 show the comparison between the power consumption for system with TSM without any restrictions and TSM with different latency limitations. As in the case with buffer limitations, (28) shows that with the increase of $N_s^*$, the time portion of working in State 2 can come closer to the maximum possible value without restrictions, defined by (18). Instead, low values of $N_s^*$ mean that the UE will work more in State 3, and there is a smaller gain in power consumption savings compared to the case without any restrictions. For $N_s^* = 2$ a significant gain can be obtained for latencies bigger than 5 s. On one hand, this is caused by the fact that limitation on latency allows for UE to stay in State 2 long enough to achieve the decrease of power consumption. At the same time, due to a reasonably high value of $N_s^*$, the UE can also move to State 3 and come back to State 2 after transmission often enough. So, generally speaking for the case with latency restrictions, as well as for the case with buffer limitations, the resulting gain in power consumption savings depends on the limitations on $N_s^*$.

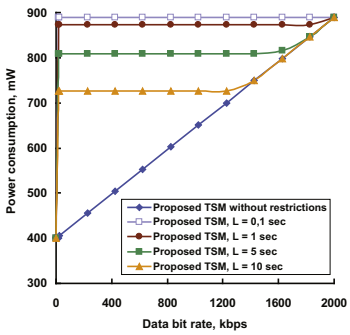



**Fig. 10.** Power consumption with latency restrictions and $N_s^* = 0.4$.

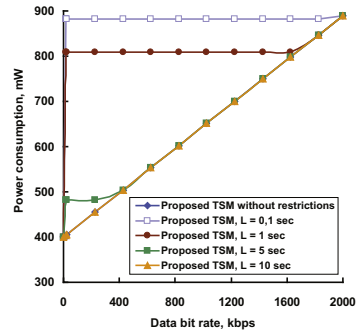



**Fig. 11.** Power consumption with latency restrictions and $N_s^* = 2$.

## 6. Energy efficient constant bit rate video transmission over 3GPP networks

In this section the analysis of video transmission over 3G networks is presented. We are focusing our research on power consumption of constant bit rate video only. One source of constant bit rate video data is a single-layer codec with rate control mechanism that helps to achieve constant bit rate of the compressed video data [36,37]. Constant bit rate transmission of variable bit rate stored video may be performed e.g. with the optimal choice of buffer size [38].

The system in general can operate as described on the Fig. 12. The UE decides the power saving mode according to its needs, and based on it the power saving controller chooses the optimal video bit rate, as demonstrated later in this section. At the same time, power saving controller finds the optimal parameters for RRC state machine, as described in Section 5. Based on the given video bit rate, the video rate controller defines the compression parameters, which are used for encoding the video source. After this the compressed bit stream is ready for the transmission with optimal TSM parameters, providing the necessary level of power consumption for the UE.

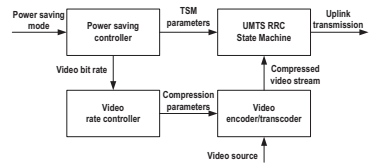



**Fig. 12.** UE power saving based on trade off between video quality and power consumption.

Another approach considers the use of scalable video coding (e.g. Scalable Extension of H.264/AVC [39]), with further extraction of the necessary amount of data under bit rate constraints (at the same time, possibly maximizing quality) [40]. The latter approach can practically be used in hybrid video surveillance systems, consisting of a camera that records video data in a scalable format and a transmission terminal that is used for the delivery of the encoded video data. For power saving reasons, in case of no motion detection, the camera can transmit the base layer only. When the motion is detected, the camera starts transmitting several (up to all) quality layers of the video sequence, thus increasing the power consumption of the mobile terminal.

In this section we evaluate the necessary amount of power for the video transmission for different levels of video quality. For practical experiments the H.264/AVC scheme in reference software JM codec v.16.2 [41] was used. Experimental results were obtained for several test video sequences: "Hall monitor", "Foreman", "Soccer" and "Akiyo" at QCIF (176 × 144) resolution, 15 Hz, 300 frames and CIF (352 × 288) resolution, 30 Hz, 300 frames, and also "Crew" at QCIF, CIF and SD (704 × 576, 30 Hz, 300 frames) resolutions. The following GOP structure was used: IPPP, GOP size = 16.

Using the JM codec, we obtained the rate-distortion performance on the test video sequences, shown on Figs. 13 and 14. Using the analysis, provided in Sections 4 and 5, it is possible to evaluate the power consumption for uplink transmission taking different restrictions into account, as it was done, for example, on Figs. 8–11. Thus, we know the dependency between the power consumption and the transmission rate. Combining these results and the rate-distortion performance, results demonstrating the trade-off between video quality and power consumption can be obtained.

We used Figs. 13 and 14 and the power consumption for TSM with proposed parameters without any restrictions (Fig. 5) to obtain Figs. 15–17. Similar results can be obtained if using power consumption for TSM with proposed parameters used for Fig. 6. They illustrate that, due to the combination of the proposed TSM parameter selection with a video encoder providing constant bit rate, it is possible to adapt the level of the power consumption based on the requirements.

Fig. 13 shows the PSNR of the luminance component vs. bit rate for four standard CIF test video sequences. Fig. 14 shows the PSNR of the luminance component vs. bit rate for one test sequence

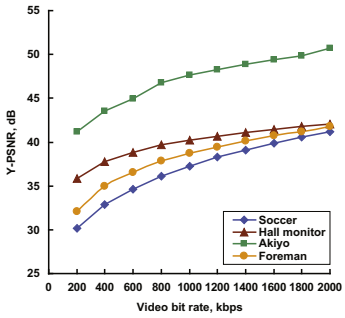



**Fig. 13.** Rate-distortion performance for CIF test video sequences.

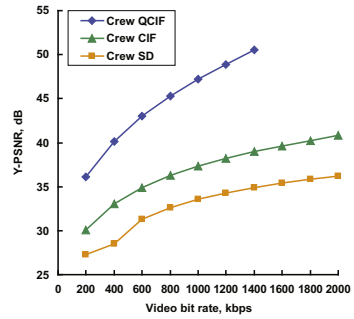



**Fig. 14.** Rate-distortion performance for "Crew" test video sequence.

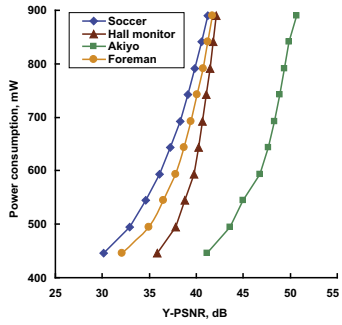



**Fig. 15.** Trade-off between video quality and power consumption for CIF test video sequences.

"Crew" in three different resolutions. As the content/resolution of these sequences varies (e.g. slow motion for "Akiyo" and fast motion for "Soccer", or, for example, QCIF and CIF resolutions for "Crew"), the performance of the video codec is also different. As a consequence, the same video quality for different sequences can be obtained at different bit rates.

As illustrated on Figs. 15 and 16, more power may be needed for the transmission of one sequence with a certain quality than for another sequence with the same level of quality. For example, for a quality level around 36 dB, approximately 450 mW will be needed for "Hall monitor" compared to 590 mW for "Soccer". A similar example can be found for "Crew" sequence, when the transmission in QCIF resolution with quality around 36 dB will require 440 mW compared to approximately twice as much power for SD resolution. That means, that it is possible to control the power consumption of the UE, by defining the required level of video quality for a particular sequence.

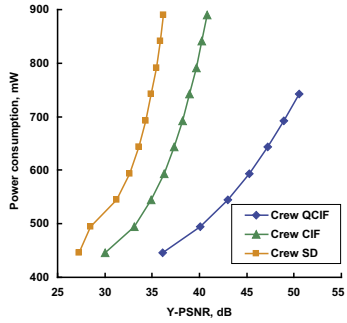



**Fig. 16.** Trade-off between video quality and power consumption for "Crew" test video sequence.

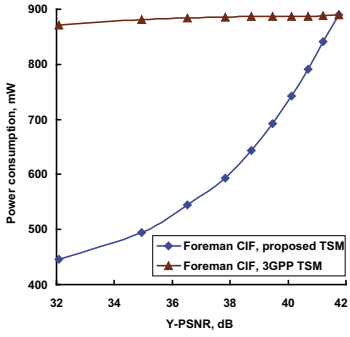



**Fig. 17.** Trade-off between video quality and power consumption for 3GPP and proposed transition state machines.

In fact, for some applications it would make sense to define the range of desired or acceptable quality (e.g. 30–40 dB for video transmission [42]), so that no additional power is spent for the unnecessary improvement of the video quality. Thus, for the provided data on Figs. 13 and 15, it could be beneficial to restrict the quality for "Akiyo" to say 45 dB, and, in case of transmission of several sequences, spend the saved power on other sequences. So, we suggest that a cross-layer strategy to optimize the power and quality performances should be sequence-dependent.

As we consider the case of restricted power consumption, we try to maximize quality within these constraints. Fig. 17 shows the gain in the video quality vs. power consumption for the 3GPP and proposed TSMs for "Foreman" video sequence in CIF resolution. For the 3GPP TSM it is possible to achieve only a small gain in power consumption savings even for a significant decrease in a bit rate. The proposed state machine uses the information about the bit rate and optimizes the parameters, as was described in Section 5, so that a significant gain in power savings is achieved (up to two times for low PSNR values for the sequence "Foreman"). The proposed approach can be used for adaptive power control for the mobile device. Depending on the power resources, the transmitting device can decrease/increase the power consumption by varying the quality of the transmitted video data. For example, if it is necessary to have the high level of power savings or the transmission will last a long time, a mobile device can switch to the energy saving mode. But if there is no need for power savings on the device in the current moment, it can work with the highest quality of the video data.

## 7. Conclusion

In this paper we have considered the power consumption of data transmission over 3G networks for smartphones. The first part of the analysis is related to the states in RRC in general. We have proposed a power consumption model based on our experimental results on a smartphone Nokia N900, taking into account packet size and transmission intervals. Our model provides a better approximation to the experimental results than the referenced model based on data rate. The second part of the paper discussed the uplink power consumption analysis of the transition state machine. We proposed a method for parameter selection for the 3GPP transition state machine that allows to significantly decrease the uplink power consumption on the mobile device taking signaling traffic, buffer size and latency restrictions into account. Furthermore, we presented an analysis of power consumption for constant bit rate video transmission in 3G networks. We achieved a gain in power consumption vs. PSNR for transmitted video of our proposed method over the one used currently in 3GPP. Our results demonstrate that, depending on the requirements for power consumption, a mobile device can balance the video quality and transmission power, permitting to perform adaptive power management. Moreover, this work can be further extended to the LTE systems, as also there a TSM exists with a low latency for transition between idle and connected states [43]. Thus, a similar approach can be applied for energy savings in LTE.

## Acknowledgments

This work was partially supported by the ECEWA project and by the Academy of Finland (Project No. 213462, Finnish Programme for Centres of Excellence in Research 2006–2011).

## References

[1] The world in 2011, ICT facts and figures, International Telecommunication Union, 2011. Available on: <www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf>.
[2] T. Stockhammer, T. Oelbaum, T. Wiegand, H.26L video transmisson in 3G wireless environments, in: Proceedings of International Conference on Third Generation Wireless and Beyond (3Gwireless'02), 2002.
[3] Cisco visual networking index: global mobile data traffic forecast update, 2010–1015, 2011. Available on: <www.cisco.com>.
[4] N. Ravi, J. Scott, L. Han, L. Iftode, Context-aware battery management for mobile phones, in: Proceedings of the Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PERCOM'08), 2008.
[5] G. Perrucci, Energy Saving Strategies on Mobile Devices, Ph.D. Thesis, 2009.
[6] K. Pentikousis, In search of energy-efficient mobile networking, IEEE Communication Magazine (2010).
[7] J. Zhang, D. Wu, S. Ci, H. Wang, A. Katsaggelos, Power-aware mobile multimedia: a survey, Journal of Communications 4 (9) (2009).
[8] K. Christensen, C. Gunaratne, B. Nordman, A. George, The next frontier for communications networks: power management, Computer Communications Journal 27 (2004).

[9] J.-H. Yeh, C.-C. Lee, J.-C. Chen, Performance analysis of energy consumption in 3GPP networks, in: Proceedings of Wireless Telecommunications, Symposium, 2004.
[10] N. Sklavos, K. Touliou, A system-level analysis of power consumption and optimizations in 3G mobile devices, in: Proceedings of the 1st International Conference on NewTechnologies, Mobility and Security (NTMS'07), 2007.
[11] J. Adams, G. Muntean, Power save adaptation algorithm for multimedia streaming to mobile devices, in: Proceedings of IEEE International Conference on Portable Information Devices, 2007.
[12] J. Korhonen, Y. Wang, Power-efficient streaming for mobile terminals, in: Proceedings of the International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'05), 2005.
[13] G. Perrucci, F. Fitzek, G. Sasso, M. Katz, Energy saving strategies for mobile devices using wake-up signals, in: Proceedings of 4th International Mobile Multimedia Communications Conference (MobiMedia 2008), 2008.
[14] B. Yu, L. Wang, J. Manner, Energy-efficient web access on mobile devices, in: Proceedings of IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social, Computing, 2010.
[15] STMicroelectronics VS6451 (ultra small CIF+ reflowable camera module) product description. Available on: <www.st.com>.
[16] H.-C. Chang, J.-W. Chen, B.-T. Wu, C.-L. Su, J.-S. Wang, J.-I. Guo, A dynamic quality-adjustable H.264 video encoder for power-aware video applications, IEEE Transactions on Circuits and Systems for Video Technology 19 (12) (2009).
[17] Q. Zhang, Z. Ji, W. Zhu, Y. Zhang, Power-minimized bit allocation for video communication over wireless channels, IEEE Transactions on Circuits and Systems for Video Technology 12 (2002).
[18] Y. Eisenberg, C.E. Luna, T. Pappas, R. Berry, A.K. Katsaggelos, Joint source coding and transmission power management for energy efficient wireless video communications, IEEE Transactions on Circuits and Systems for Video Technology 12 (2002).
[19] Y. Chen, S. Zhang, S. Xu, G.Y. Li, Fundamental trade-offs on green wireless networks, IEEE Communications Magazine (2011).
[20] M. Kim, J. Choi, M. Kang, Trade-off guidelines for power management mechanism in the IEEE 802.16e MAC, Computer Communications Journal 31 (2008).
[21] S. De Vuyst, K. De Turck, D. Fiems, S. Wittevrongel, H. Bruneel, Delay versus energy consumption of the IEEE 802.16e sleep-mode mechanism, IEEE Transactions on Wireless Communications 8 (11) (2009).
[22] E. Belyaev, T. Koski, J. Paavola, A. Turlikov, A. Ukhanova, Adaptive power saving on the receiver side in digital video broadcasting systems based on progressive video codecs, in: Proceedings of the 11th International Symposium on Wireless Personal Multimedia, Communications, 2008.
[23] E. Belyaev, V. Grinko, A. Ukhanova, Power saving control for the mobile DVB-H receivers based on H.264/SVC standard, in: Proceedings of the 8th Wireless Telecommunication, Symposium, 2008.
[24] F. Qian, Z. Wang, A. Gerber, Z.M. Mao, S. Sen, O. Spatscheck, Characterizing radio resource allocation for 3G networks, in: Proceedings of the 10th Annual Conference on Internet, Measurement, 2010.
[25] H. Haverinen, J. Siren, P. Eronen, Energy consumption of always-on applications in WCDMA networks, in: Proceedings of the 65th Semi-Annual IEEE Vehicular Technology Conference, 2007.
[26] White paper from wavion wireless networks, metro zone Wi-Fi for cellular data offloading, 2010. Available on: <www.wavionnetworks.com>.
[27] K. Mahmud, M. Inoue, H. Murakami, M. Hasegawa, H. Morikawa, Measurement and usage of power consumption parameters of wireless interfaces in energy-aware multi-service mobile terminals, in: Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio, Communications, 2004.
[28] K.-H. Lee, J.-H. Park, J.-S. Koh, User experience analysis of smartphone web surfing in UMTS networks, in: Proceedings of the 72th IEEE Vehicular Technology Conference, 2010.
[29] C. Johnson, Radio Access Networks for UMTS: Principles and Practice, John Wiley & Sons, Ltd., 2008.
[30] P.H.J. Perala, A. Barbuzzi, G. Boggia, Theory and practice of RRC state transitions in UMTS networks, in: Proceedings of 5th IEEE Broadband Wireless Access Workshop co-located with IEEE Globecom, BW-WAWS, 2009.
[31] P. Willars, Smartphone traffic impact on battery and network. Available on: <www.labs.ericsson.com/developer-community/blog/smartphone-traffic-impact-battery-and-networks>.
[32] Signals research group, smartphones and a 3G network. Available on: <www.signalsresearch.com>.
[33] T. Stockhammer, M.M. Hannuksela, T. Wiegand, H.264/AVC in wireless environments, IEEE Transactions on Circuits and Systems for Video Technology 13 (2003).
[34] F. Fitzek, M. Krishnam, M. Reisslein, Providing application-level QoS in 3G/4G wireless systems: a comprehensive framework based on multi-rate CDMA, IEEE Wireless Communications 9 (2) (2002).
[35] N. Baghaei, R. Hunt, Review of quality of service performance in wireless LANs and 3G multimedia application services, Computer Communications Journal 27 (2004).
[36] MPEG-2 TM5 rate control and quantization control. Available on: http:<www.mpeg.org/MPEG/MSSG/tm5/Ch10/Ch10.html>.
[37] I. Shin, Y. Lee, H. Park, Rate control using linear rate-$\rho$ model for H.264, Signal Processing: Image Communication 19 (4) (2004).
[38] S. Sen, J.K. Dey, J.F. Kurose, J.A. Stankovic, D. Towsley, Streaming CBR transmission of VBR stored video, in: Proceedings SPIE Symposium Voice Video and Data Communications, 1997.
[39] H. Schwarz, D. Marpe, T. Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard, IEEE Transactions on Circuits and Systems for Video Technology 17 (9) (2007).
[40] E. Maani, A.K. Katsaggelos, Optimized bit extraction using distortion modeling in the scalable extension of H.264/AVC, IEEE Transactions on Image Processing 18 (9) (2009).
[41] H.264/AVC JM reference software. Available on: <iphome.hhi.de/suehring/tml>.
[42] V. Vassiliou, P. Antoniou, I. Giannakou, A. Pitsillides, Requirements for the transmission of streaming video in mobile wireless networks, in: Proceedings ICANN, 2006.
[43] Motorola, technical white paper, long term evolution (LTE): a technical overview, 2007.

# Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band

**Alexander Lebedev,[1,*] Tien Thang Pham,[1] Marta Beltrán,[2] Xianbin Yu,[1] Anna Ukhanova,[1] Roberto Llorente,[2] Idelfonso Tafur Monroy,[1] and Søren Forchhammer[1]**

[1]*DTU Fotonik, Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[2]*Valencia Nanophotonics Technology Center, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain*
*alele@fotonik.dtu.dk

**Abstract:** The paper addresses the problem of distribution of high-definition video over fiber-wireless networks. The physical layer architecture with the low complexity envelope detection solution is investigated. We present both experimental studies and simulation of high quality high-definition compressed video transmission over 60 GHz fiber-wireless link. Using advanced video coding we satisfy low complexity and low delay constraints, meanwhile preserving the superb video quality after significantly extended wireless distance.



**OCIS codes:** (060.2330) Fiber optics communications; (060.5625) Radio frequency photonics.

**References and links**

1. M. Beltrán, J. B. Jensen, X. Yu, R. Llorente, and I. T. Monroy, "Experimental performance comparison of 60 GHz DCM OFDM and impulse BPSK ultra-wideband with combined optical fibre and wireless transmission," in ECOC, 1465–1467 (2010).
2. Z. Jia, H.-C. Chien, Y.-T. Hsueh, A. Chowdhury, J. Yu, and G.-K. Chang, "Wireless HD services over optical access systems: Transmission, networking, and demonstration," in OFC, 1–5 (2009).
3. M. Weiß, "60 GHz photonic millimeter-wave communication systems," thesis (2010).
4. A. Belogolovy, E. Belyaev, A. Sergeev, and A. Turlikov, "Video compression for wireless transmission: reducing the power consumption of the WPAN hi-speed systems," NEW2AN/ruSMART 2009, LNCS 5764, 313–322 (2009).
5. http://iphome.hhi.de/suehring/tml/.
6. T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," IEEE Trans.Circuits Syst. Video Technol. **13**(7), 657–673 (2003).
7. I. E. Richardson, *The H.264 Advanced Video Compression Standard* (Wiley, 2010).
8. http://www.vpiphotonics.com/.
9. S.-K. Yong, *60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice* (Wiley, 2011), Chap. 2.
10. S. K. Yong and C.-C. Chong, "An overview of multigigabit wireless through millimeter wave technology: potentials and technical challenges," EURASIP J. Wireless Commun. Netw. **2007**(1), 078907 (2007).
11. K.-C. Huang and D. J. Edwards, *Millimetre Wave Antennas for Gigabit Wireless Communications: A Practical Guide to Design and Analysis in a System Context* (Wiley, 2008).

## 1. Introduction

The motivations for this work are three-fold. First, the unprecedented frequency range around 60 GHz (from 4 to 9 GHz within 57-66 GHz) has been regulated for unlicensed use in a number of countries around the world. Second motivation is the introduction of high quality video services such as high-definition (HD) video conferencing and distributed video gaming. These services define both the demand for increased data rates in the access networks and need for optimization of video compression schemes. Third, efficient convergence of wired and wireless technologies is required to enable the concept of "anytime anywhere" wireless

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011
(C) 2011 OSA 12 December 2011 / Vol. 19, No. 26 / OPTICS EXPRESS B895

A. Lebedev, T. Pham, M. Beltrán, X. Yu, A. Ukhanova, R. Llorente, I. Monroy, S. Forchhammer, "Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band", *Optics Express Journal*, Vol. 19, Iss. 26, 2011.

connectivity. Radio-over-fiber (RoF) is considered a promising example of such integration for optical networks [1].

Previous research in the area of 60 GHz RoF video transmission suggests the use of uncompressed video [2,3]. The main drawback of this approach is reduced flexibility in terms of bitrate: bitrates are fixed depending on resolution, number of bits per pixel, and frame rate of the video sequence. This therefore results in extremely complex adaptation of the HD video system to significant signal-to-noise ratio (SNR) drops caused by either severe shadowing in non-line-of-sight (NLOS) case or extremely high attenuation – problems that are typical for 60 GHz systems.

Source coding (compression) gives us desirable flexibility of bitrate but at the expense of introducing delay and increase of power consumption. However, there is a trade-off between the power needed to radiate larger bandwidth for uncompressed video and the power consumed for the computations of an encoder and a decoder for compressed video transmission. According to [4], low complexity compression can, in fact, bring about reduction in power consumption for a 60 GHz wireless video transmission system compared to the uncompressed case, while at the same time keeping delay under the acceptable limit.

In this work we explore the notion of joint optimization of physical layer parameters of a RoF link (power levels, distance) and the codec parameters (quantization, error-resilience tools) based on peak signal-to-noise ratio (PSNR) as an objective video quality metric. We experimentally demonstrate, first time to our knowledge, the combined optical access and wireless transmission of compressed HD video in the 60 GHz band employing simple envelope detection technique.

**2. Experimental setup**

The experimental setup of the 60 GHz optical-wireless RoF system is shown in Fig. 1. The binary sequence corresponding to compressed video file was uploaded in an arbitrary waveform generator (AWG). The non-return-to-zero (NRZ) electrical signal on the output of the AWG directly modulated a 1550 nm laser. After the baseband data modulation, frequency up-conversion to the 60 GHz band was performed by driving a Mach-Zehnder modulator (MZM) biased at the minimum transmission point with a 30 GHz sinusoidal signal. A polarization controller (PC) was used before the MZM to minimize its polarization-dependant losses. After the MZM, two sidebands with a frequency spacing of $2f_{LO}$ were generated according to the double sideband-suppressed carrier (DSB-SC) intensity modulation scheme (see Fig. 2). Optical carrier suppression of approximately 13.6 dB is achieved limited by the MZM extinction ratio. The generated sidebands have the same optical power and the locked phase. Subsequently, an Erbium doped fiber amplifier (EDFA) is employed to compensate the losses, and an optical band pass filter (OBPF) is used afterwards to mitigate the amplified spontaneous emission (ASE) noise produced by the EDFA. Then the signal is launched into a 20 km span of non-zero dispersion shifted fiber (NZDSF). We employ the NZDSF in order to minimize dispersion induced impairments. A variable optical attenuator (VOA) is employed to control the optical power impinging the photodiode (PD) in order to evaluate BER performance of the system as a function of the received optical power.

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

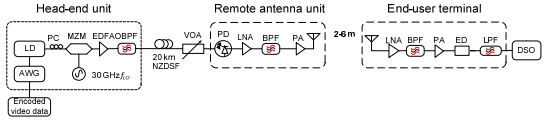



Fig. 1. Experimental 60 GHz optical-wireless RoF system with envelope detection, LD-laser diode, PC-polarization controller, MZM-Mach-Zehnder modulator, LO-local oscillator, EDFA-Erbium doped fiber amplifier, OBPF-optical band pass filter, PD-photodiode, LNA-low noise amplifier, PA-power amplifier, BPF-band pass filter, ED- envelope detector, LPF-low pass filter, DSO-digital sampling oscilloscope.

After photodetection the 60 GHz signal was amplified (gain of amplifiers – 16 dB and 28.7 dB) and filtered (58.1-61.9 GHz) before feeding it to an antenna for up to 6 meters of wireless transmission. After receiving the signal with an antenna and following filtering (58.1-61.9 GHz) and amplification (gain of amplifiers – 16 dB and 28.7 dB) envelope detection was employed for down-conversion. The detected envelope is low-pass filtered and digitized by a digital sampling oscilloscope (DSO). Both the transmitting and receiving antennas used throughout the experiment are commercially available horn antennas with 20 dBi gain and 12° beam width. Bitrates that were transmitted over the fiber are low compared to similar research setups. This explains a good performance to a certain extent, but we emphasize that reduction of bitrate does not lead to a significant video quality deterioration.

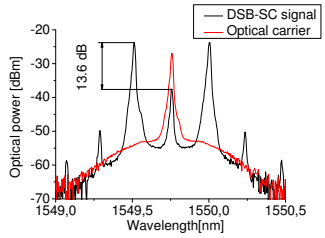



Fig. 2. Optical spectra on the input of the PD.

The encoding was performed using the Joint Model (JM) 17.0 reference software implementation of the H.264/Advanced Video Codec (AVC) [5]. It is a realistic scenario since H.264/AVC is one of the latest industrial video coding standards covering a wide range of applications, including, coding for transmission over wireless links and HDTV coding [6]. An Intra coding mode only and a frame slicing mechanism were employed to achieve the low delay requirement. Both mechanisms are improving the error-resilience as well [7]. Slicing was performed with the use of flexible macroblock ordering (FMO).

H.264 is not capable of coping with single-bit errors: its mechanisms of error-resilience on the encoding side and error concealment on the decoding side are adjusted to cope with packet loss when the packets affected by the errors are discarded such as usually occurs in networks. Packet error rate (PER) depends on the bit error rate (BER) and the size of the packet; in general, the noisier the transmission the shorter the length of the packet is desirable. Initially in the experiment we used the packet size equal to 2500 bytes, each packet containing a slice of the frame; afterwards we have been using packets of length of 3000 and 3500 bytes for

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

simulation. The uncompressed HD test video sequence 'blue sky' was used for encoding and transmission. The sequence was originally shot in 4:2:0 format 8 bits per color 1920 × 1080 pixels. However, in order to model the most bitrate demanding case upsampling to 4:4:4 format was performed (uncompressed bitrate – 3 Gbps for the frame rate of 60 frames per second).

We use PSNR as an objective quality metric for video, which is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2, \tag{1}$$

$$PSNR = 10\log_{10}(\frac{L^2}{MSE}), \tag{2}$$

where MSE stands for mean squared error, N is the number of pixels in the image or video signal, and $x_i$ and $y_i$ are the i-th pixel values in the original and the distorted signals, respectively. L is the dynamic range of the pixel values. For an 8 bits/pixel signal, L is equal to 255. PSNR is evaluated for the luminance component of the transmitted video signal.

**3. Composite fiber-wireless channel modeling for 60 GHz band**

The difficulty of the modeling arises from the fact that we need to account for both the impairments induced by the wireless and the fiber-optic channels. We performed the modeling of the fiber-optic channel with VPI software [8]. The wireless channel model was implemented in Matlab and combined with VPI channel model afterwards. We combine below the description of the channel model with the excerpts from experimental measurements that allow us to simplify the model.

Noise processes in the optical part of the setup (such as amplified spontaneous emission (ASE) noise, Johnson noise, shot noise at the photodiode), attenuation and dispersion in the fiber are simulated in VPI software. We set the numerical values for these parameters according to the specifications of equipment we used in the experimental setup.

We performed the modeling of the wireless channel according to the physical parameters of the devices that have been used in the scheme and references on theoretical parameters taken from [9–11].

The path loss (attenuation) at 60 GHz is much more severe than the path loss at the frequencies that are currently employed for Wireless Personal Area Networks (WPAN). Theoretical description for this phenomenon is provided by Friis formula [9], according to which attenuation in the air is proportional to the frequency squared. It is known that the line-of-sight (LOS) attenuation of the 60 GHz wireless channel can be modeled with a log-normal model [11]. Parameters for this model have been defined through the extensive measurements presented in a number of publications. Summary on the parameters for different experimental environments can be found in [9]. We perform the modeling of the system without taking into account frequency dependency of the path loss. To the best of our knowledge, frequency dependent models for 60 GHz system have not yet been reported.

Influence of the noise on the signal can be modeled with the following formula [10,11]:

$$\begin{aligned} SNR = {} & P_{tx} + G_T + G_R + G_{LNA_{Rx}+PA_{Rx}} + G_{LNA_{Tx}+PA_{Tx}} - PL - \\ & -(10\log_{10}(KTB) + NF_{LNA_{Rx}+PA_{Rx}} + NF_{LNA_{Tx}+PA_{Tx}}), \end{aligned} \tag{3}$$

$$PL = PL(d_0) + 10n\log_{10}(\frac{d}{d_0}), \tag{4}$$

where $P_{tx}$ in our case is the RF power on the output of the PD, $G_T$ and $G_R$ are the gain of transmitting and receiving antennas respectively, $G_{LNA_{Rx}+PA_{Rx}}$ and $G_{LNA_{Tx}+PA_{Tx}}$ are gains of

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

amplification cascades at transmitting and receiving parts of the scheme respectively, PL is the distance-dependent path loss (attenuation) in the air. The terms in brackets represent noise contributions. The first term represents the Johnson noise, second and third represents noise contributions from amplifiers. Parameters $d_0$ and $d$ in Eq. (4) represent the reference distance (we used 1 m according to [9]) and the distance between a transmitter and a receiver respectively, $n$ denotes path loss exponent.

The formula does not account for shadowing caused by LOS obstruction, but this resembles the experimental setting where we were working with the LOS scenario only.

Phase noise modeling for the channel was excluded after the experimental examination of the phase noise of the oscillator presented in Fig. 3. Figure 3 shows the high quality of the electrical oscillator for 3 cases: measuring the phase noise of LO, setup without fiber transmission up to a transmitting antenna (optical back-to-back) and after 20 km of NZDSF. Figure 3 also illustrates the fact that contribution from the system to the phase noise is insignificant. Moreover, it could be excluded from consideration, because after wireless transmission we finally recover with ED only the amplitude of the signal, and therefore discard information about phase or frequency.

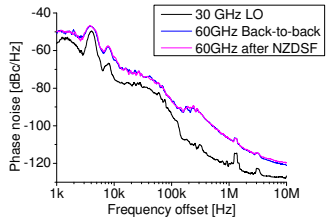



Fig. 3. Phase noise of RF subcarriers.

The model does not take into account the nonlinear effects that are reported for Power Amplifiers in [9]. Nevertheless we regard the model as feasible since the power after the PD is low, so we work in the linear region. Indeed, the power on the output of the PA at the transmitting side given the power at the photodiode of –10 dBm is around –6 dBm. Typically nonlinear effects are observed in the region above 0 dBm [9]. The RF-spectrum measured is depicted in Fig. 4. We refer to the power before the antenna, as the power before radiation $P_{br}$. Therefore the equation for wireless channel simulation based on Eq. (3) and Eq. (4) could be transformed into:

$$SNR = P_{br} + G_T + G_R + G_{LNA_{Rx}+PA_{Rx}} - PL(d_0) - 10n\log_{10}(\frac{d}{d_0}) - \\ -(10\log_{10}(KTB) + NF_{LNA_{Tx}+PA_{Tx}} + NF_{LNA_{Rx}+PA_{Rx}}). \quad (5)$$

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

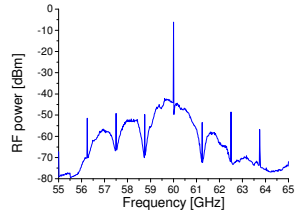



Fig. 4. RF spectrum measured before the antenna.

Typical parameters for the path loss at the reference distance and the path-loss exponent has been found in the references [9,10]. Values for the parameters that are used in modeling of the wireless channel are listed in the Table 1 below.

**Table 1. System Parameters for Modeling**

| Parameter | Numerical value |
|---|---|
| Center frequency, GHz | 60 |
| Joint noise figure Tx amplifiers (LNA + PA), dB | (6 + 7) = 13 |
| Joint noise figure Rx amplifiers (LNA + PA), dB | (6 + 7) = 13 |
| Joint gain of Tx amplifiers (PA + LNA), dB | (28.7 + 16) = 44.7 |
| Joint gain of Rx amplifiers (PA + LNA), dB | (28.7 + 16) = 44.7 |
| Gain of the Tx antenna, dBi | 20 |
| Gain of the Rx antenna, dBi | 20 |
| Bit rate, Mbps | 312.5/1250 |
| Distance, m | 2-6 |
| Reference path loss at 1 meter, dB | 57.5 |
| Path loss exponent | 1.77 |
| Ambient temperature for Johnson noise modeling, K | 298 |

We perform attenuation of the signal and addition of the Additive White Gaussian Noise (AWGN) in VPI, the noise power and attenuation to achieve SNR described in Eq. (5) are calculated in Matlab.

**4. Results and discussion**

Our goal for optimization is to achieve the best video delivery quality for a given link budget. With regards to the role of the quantization of transform coefficients of the coded video in the optimization, roughly speaking, the smaller the quantization parameter size, the smaller the source distortion (loss due to compression), but the larger the channel distortion it may cause. In the experiment we explored two cases. First, the chosen test video sequence ('blue sky' 4:4:4) was encoded with bitrate of 312.5 Mbps. Second, the tested video sequence was encoded in a high quality setting with the quantization parameter equal to 1, which gave us a compression ratio of 3.

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

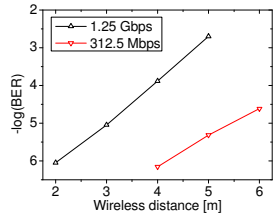



Fig. 5. BER as a function of the wireless distance.

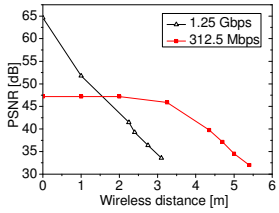



Fig. 6. PSNR as a function of the wireless distance.

On the Fig. 5 BER at the power level at the photodiode equal to –10 dBm as a function of the wireless distance is depicted. From the Fig. 5 we can see that in general the distortion induced by the wireless channel is severe in our system, but video coded with the use of higher quantization parameter has greater dynamic range of wireless distance, as shown in Fig. 6. The distance equal to 0 corresponds to the distortion introduced by the compression only. When we increase the wireless distance, in the beginning, the source distortion is dominant, and the use of lower quantization parameter is reasonable. Anyhow, we lose the advantage of lower distortion after around 2 m of transmission when video is evaluated based on the PSNR metric only. This shows the potential of optimization of the power budget of the system under the constraint of video quality. We obtain similar curves for changing optical power level at the photodiode at 5 m of wireless distance, as shown in Fig. 7 and Fig. 8. With the higher video compression we can work at lower optical power levels. At the same time, we should note that the video quality is high in both cases, and deterioration induced by the compression itself can be regarded as non-significant (PSNR of the video unimpaired by the channel is higher than 45 dB in both cases).

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

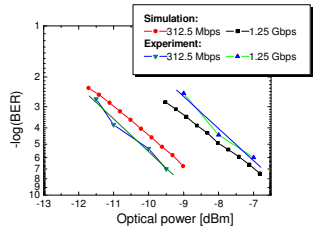



Fig. 7. BER as a function of the optical power at the photodiode.

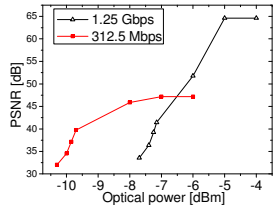



Fig. 8. PSNR as a function of the optical power at the photodiode.

Curves provided by simulation for 5 meters distance and dynamically changing optical power levels show close resemblance that verifies the correctness of the simplified wireless channel model employed. We do not provide simulation-based curves for PSNR, because our simulation is based on the analytical estimation of BERs with the use of VPI software, and we therefore do not have traces including erroneous bits to analyze video performance.

**5. Video coding for 60 GHz radio-over-fiber**

We employed video coding parameters in a simplified setting that is suitable for both conferencing applications and distributed video gaming. The main constraints for such type of an application are delay and energy consumption. As a part of simplified setting we were using Universal Variable Length Coding (UVLC) for entropy coding that is considered a lower complexity solution [7]. All coding experiments were performed in intra mode thus eliminating the need for long buffering time, and satisfying low delay requirement. The simulation below was performed with bit traces including erroneous bits.

H.264/AVC encoder employs the number of error-resilience tools: slicing of the frame, data partitioning, arbitrary slice ordering, and redundant coded slices [7]. Below we present simulation on two major tools providing error resilience: slices and Flexible Macroblock Ordering (FMO). On the decoder side, there are two error concealment tools used in JM 17.0 reference software implementation of H.264/AVC codec, one exploiting spatial information only, suitable for intra frames (the one used in the experiment and simulation), and one

#155834 - $15.00 USD
Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

exploiting temporal information. Details on the error concealment algorithms used can be found in [5].

First we performed the simulation with a different size of the packet (each containing one slice of the frame). Employing the smaller slices enables us to receive acceptable video quality in the regions with higher BER, and therefore extends distance for acceptable quality of video transmission. Indeed, enabling packets of shorter length reduces the amount of information lost when the packet is discarded, enabling decoder to reconstruct impaired parts of the picture better from unimpaired blocks of neighboring pixels. The simulation results are illustrated on Fig. 9.

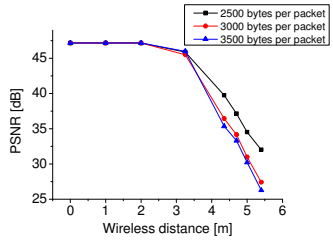



Fig. 9. PSNR as a function of the wireless distance for different packet sizes of the encoded video for the bitrate of 312.5 Mbps.

Below we also present the simulation results for enabling FMO in H.264 reference software [5]. H.264/AVC is the first standard defining this error-resilience tool [7]. In case if we do not use FMO, the images will be composed of a single slice groups with the macroblocks in a scan order. If we employ this algorithm, then when we lose a slice of the video frame, we can make better approximation with the neighboring blocks and therefore, presumably, can achieve gain in PSNR. Results of the simulation for the packet size of 3000 bytes are depicted in the Fig. 10. FMO shows up to 3 dB improvement of PSNR.

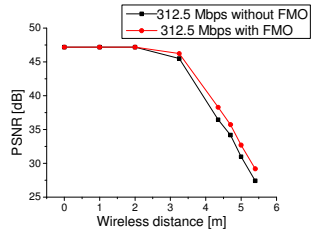



Fig. 10. PSNR performance as a function of the wireless distance for FMO effect estimation.

Coding simulations show the effect of employed source coding error-resilience mechanisms for a particular simplified setting of H.264/AVC and 60 GHz RoF setup as an example of physical layer architecture suitable for transmission high quality HD video

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

content. Employed tools of H.264/AVC show the greater robustness of video provided by advanced video coding against impairments induced by 60 GHz fiber-wireless channel.

**6. Conclusions**

Our experiment and simulation demonstrates the trade-off between the distortion introduced by the source (lossy compression) and distortion introduced by channel for high quality HD video transmission over 60 GHz RoF fiber-wireless links. We have achieved significant extension of wireless distance employing low complexity physical layer solution for detection of RF modulated signal. Our work demonstrates the solutions for improving robustness and reach of simplified converged fiber-wireless RoF communication links provided by advanced video coding.

**Acknowledgments**

This work has been partly funded by the European Commission under FP7 ICT-249142 FIVER project and by the by the Spanish Ministry of Science and Innovation under the TEC2009-14250 ULTRADEF project.

#155834 - $15.00 USD Received 3 Oct 2011; accepted 28 Nov 2011; published 8 Dec 2011

# Extending JPEG-LS for low-complexity scalable video coding

Anna Ukhanova[a], Anton Sergeev[b] and Søren Forchhammer[a]

[a]DTU Fotonik, Kgs. Lyngby, Denmark;
[b]SUAI, St. Petersburg, Russia

**ABSTRACT**

JPEG-LS, the well-known international standard for lossless and near-lossless image compression, was originally designed for non-scalable applications. In this paper we propose a scalable modification of JPEG-LS and compare it with the leading image and video coding standards JPEG2000 and H.264/SVC intra for low-complexity constraints of some wireless video applications including graphics.

**Keywords:** scalable JPEG-LS, H.264/SVC, JPEG2000, wireless video transmission

## 1. INTRODUCTION

JPEG-LS[1] was designed as an efficient algorithm for completely lossless compression of still images and controlled lossy mode where a precise upper bound on the maximal error in pixel value could be predefined by the user. The main advantages of these algorithms are very low computational complexity, perfect coding efficiency at high rates and possibility of near-lossless compression.[2,3] All this allows JPEG-LS to successfully compete for some applications with the state-of-the-art compression algorithms for still images (JPEG2000 [4]) and video (H.264 [5]). JPEG-LS comes to foreground for application with strong constraints for implementation complexity and memory consumptions: processing and storage of medical, airspace images, maps; mobile video; desktop graphics; transmission of HD video and etc. Unfortunately JPEG-LS has also some evident shortcomings. In contrast to its competitors, the original JPEG-LS has no means for building scalable video streams and organizing multithreaded transmission. In scalable video coding partial loss of the scalable stream does not irreparably affect the decoding process: the decoder may reconstruct image with reduced quality level (without re-compression and retransmission). The successful reception of all the compressed streams results in decoding and reconstruction of the whole image with the same quality level as at the coder side. Scalability is extremely important in modern video coding and useful in video transmission, downloading, providing variable quality access etc..[6,7] There are no such a features in the standard JPEG-LS. In this work we propose a new scalable extension of JPEG-LS and compare it with two generally recognized leaders in the area of image and video coding: JPEG2000 and H.264/SVC in intra mode. Modern full-featured image/video compression algorithms support a variety of compression modes, each with different trade-offs among efficiency of compression, loss of fidelity in the compression and the amount of computation required to encode/decode. Before designing and comparing the compression algorithms, we should determine which modes are suitable based on target application scenarios and limits. Today video transmission in wireless networks (WPAN/WMAN/WLAN) has become a hot topic in the industry.[7–9] Therefore this paper aims at providing a performance evaluation of the proposed and other algorithms for the task of real-time wireless high-quality low-complexity video transmission. Comparison parameters are selected accordingly to the requirements of wireless video applications:

- Reduced computational complexity because in mobile terminals (laptops, phones, etc) battery life and processing power are strongly limited.
- High coding efficiency at high rates. Throughput of wireless technologies increased during the last decade (11 Mbps 802.11b to 160 Mbps 802.11n in WLAN, 3Mbps of Bluetooth to 3Gbps of 802.15.3TG3 in WPAN). So looking forward we focus on high transmission rates and consequently consider low compression ratios.
- Scalability and error resilience, controlled quality reduction achieved by partial loss of bit stream.

Further author information: (Send correspondence to Anna Ukhanova)
E-mail: annuk@fotonik.dtu.dk, Telephone: +45 4525 6567

Image Processing: Algorithms and Systems IX, edited by Jaakko T. Astola, Karen O. Egiazarian,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7870, 78701D · © 2011 SPIE-IS&T
CCC code: 0277-786X/11/$18 · doi: 10.1117/12.887416



A.Ukhanova, A. Sergeev, S. Forchhammer, "Extending JPEG-LS for low-complexity scalable video coding", *IS&T/SPIE Electronic Imaging*, San Francisco (CA), USA, 2011.

- Universality i.e. effective coding of different types of images including photo, computer graphics and synthetic images.

The paper is organized as follows. Section 2 includes a short overview and basis for profile selection of JPEG-LS and competitors. In Section 3 the scalable JPEG-LS modification is proposed. Section 4 describes comparison methodology and presents comparison of the compression algorithms according to the selected requirements. Section 5 presents results of the performance comparison of the suggested scalable JPEG-LS, standard H.264 intra and JPEG 2000. Section 6 concludes the paper.

## 2. CODECS OVERVIEW

An overview of the selected codecs is given here.

### 2.1. JPEG2000 compression

JPEG2000 is the current ISO/ITU-T standard for still and motion image coding. The RCT (Reversible Color Transform) color format is used, decribed in the standard. JPEG2000 supports many interesting features such as lossless and lossy compression, multi-resolution representation, scalable and Region Of Interest (ROI) coding, tiling, blocking, error resilience and a flexible file format. But on the other hand JPEG2000 is more complex and slower than the prior and still widely-used JPEG standard.[10]

### 2.2. H.264/SVC compression

The Moving Picture Experts Group (MPEG) has introduced the Scalable Video Coding standard, which is an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard .[11] It has additional properties like scalability. So, on one hand, H.264/SVC could be used in the situation when we have many receivers and it is necessary to receive the data at different bitrates. Another case is when we have to control the transmission rate depending on the situation in the channel. Other important features of H.264 are integrated rate control, deblocking filter and error resilience.

### 2.3. JPEG-LS compression

JPEG-LS[1,12] is the international standard for lossless and near lossless still images compression. The main advantage of JPEG-LS is a possibility to set up the maximum error value per pixel by choosing a bound on the differences for near-lossless coding (so-called near parameter or lossy-factor). Another plus of JPEG-LS is an extremely low level of implementation complexity[8] and memory consumption at the encoder side. The encoding process requires less than two rows of samples only (less than 10KByte). Unfortunately the original JPEG-LS standard has no means for building scalable video streams and organize multithreaded transmission in contrast with its competitors. Scalability (see Section 4.2) here means a possibility of lossy reconstruction with smaller quality level in case of partial data loss. This feature is very important for wireless video applications because wireless channel may be quite non-stable and some data could be lost during the transmission. Therefore in the next section we propose a scalable version of the JPEG-LS to address this shortcoming.

### 2.4. Codec Profiles

During the experiments we compared different profiles and feature sets for the JPEG2000, H.264/SVC and JPEG-LS image or video compression algorithms. The following configuration sets are selected for analyzing the codecs from the point of the tradeoff between rate-distortion characteristics and complexity:

- JPEG2000: RCT colorpace transforms, tile size 1280x8, codeblock size 512x8, number of DWT levels is equal 2, reversible wavelet 5-3;
- JPEG-LS: scalable modification, per pixel processing, lossless RCT colorspace transform, no subsampling, Golomb coding. It was noted[8] that for HD picture size and 60 FPS it is possible to design and implement in RTL (150MHz clock rate, TSMC 60nm technology) the encoder with total power consumption of around 10mW;

– H.264/SVC intra: YUV colorspace, all intra prediction, 4x4 DCT, Intra/IPCM detection, CABAC. The power consumption for H.264 for similar parameters was given as well as for JPEG-LS as stated above.[8] H.264/SVC has a power consumption, which is roughly at least twice that of JPEG-LS. However, if the channel fluctuates, it is possible to use a scalable stream for power saving.[13]

## 3. SCALABLE JPEG-LS COMPRESSION

A simple scalable version of JPEG-LS has been proposed[14] by sending one or more least-significant bit-planes uncoded and only coding the reduced precision image using JPEG-LS. We consider another approach for better performance and/or finer granularity in the scalable format.

Our proposed scalable JPEG-LS-based algorithm utilizes the well-known differential coding approach.[15–17] Firstly data is compressed in lossy mode, then the main (primary) compressed stream is constructed providing the base quality level. Then the difference between an original and compressed image is calculated pixel by pixel. This difference is also encoded and forms the second (secondary) stream. The main question here is: what compression method should be used for the second stream encoding?

The main idea of the proposed scalable JPEG-LS is to compress the input image (or video intra-frame) in two (or more) steps and create two (or more) sub-streams joint in one global scalable bit-stream. At the first step JPEG-LS is used in lossy mode for encoding the original image. The compressed data forms the primary (or the first) sub-stream. The lossy mode increases the compression rate (and thereby decreases the amount of data to be transmitted) but distorts the original image. At the second step at the transmitter side a residual image is calculated as the difference between the original and the reconstructed images (after the first step). Then the residual image is compressed by lossless JPEG-LS and this constitutes the second bit stream for the wireless transmission.

Therefore the encoder generates a single global bitstream, which may include the primary bit stream and the second bit stream. The global bit stream could be transmitted over a stable or an un-stable communication wired or wireless communication channel. The codec may use a progressive approach to provide SNR scalability that allows supporting features such as multi-streaming and prioritized transmission.

A successful reception and decoding of both sub-streams provides a lossless image reconstruction at the receiver side (i.e. reconstructed image is equal the original one). At the same time if there are any problems with the second stream there is a possibility to decode only the primary stream and show the image with visually acceptable quality level (arount 42db PSNR for for a maximum pixel value error of 2). Similarly if the quality level of the primary stream is satisfactory for the end-user then the decoding of the second stream (even if it is successfully received) could be omitted.

The maximum error value per pixel (near parameter, $n$) for constructing the primary bit stream should be selected according to the requirements of the real applications. Figure 1 gives an overview of the algorithm. It is summarized by the following formulas: Consider $x_{ij}$ to be the pixel of the input image with coordinates $i$ and $j$ accordingly. This pixel is firstly compressed with the lossy factor $n_1$. After successful reconstruction of the pixel it is equal $\hat{x}_{ij}$. The corresponding difference between the original pixel and the reconstructed one, $x_{ij}$ - $\hat{x}_{ij}$, is denoted $d_{ij}$. Prior to making the additional lossless compression of the differential image, all values are shifted by the value of lossy factor $n_1$, $d'_{ij} = d_{ij} + n_1$, to get non-negative values. Then the same block of lossless compression is applied to the image ($d'_{ij}$) of restricted range. This forms the second bit stream of compressed pixels.

The bigger $n$ is selected, the smaller the size of the primary stream will be and the more stable it is to the sudden throughput fluctuations. In our investigations we selected $n = 1$ because in this case the total size of the primary and secondary bit streams is minimum while the size of both bitstreams are approximately equal. One could notice that JPEG2000 and SVC are truly scalable while the presented modification of JPEG-LS relates mainly to progressive two-step scalability.

It should be noted though that the simple two-step coding is well-correlated with the existing usage models for the end-user side:

– the picture of the ideal quality is shown while the wireless channel is in the "good" state,

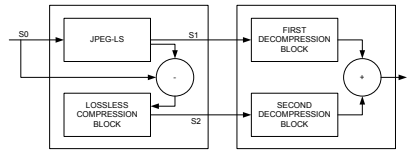



**Figure 1.** Full lossless decompression

– there is a way of the instantaneous and visually graceful, ideally imperceptible quality degradation in case of sudden noise level fluctuations.

Depending on the image type in lossless mode the compression performance of Scalable JPEG-LS is 2-23% worse then Standard JPEG-LS . With the growth of n this gap slowly increases.

Our proposal also may be extended to a nested structure (as e.g. for near-lossless coding[18]): For the JPEG-LS parameter, NEAR = $n_i$, i = 1, 2, ... we get the interval: $\prod_i(2 \times n_i + 1)$. Using $n_1$=$n_2$=1 we have intervals of 9, 3, 1(=lossless) in a scalable manner. For finer granularity we may use, the visual quantization in JPEG-LS part 2, which context adaptively chooses between near parameters $n$ and $n + 1$. This may readily be treated as above based on the larger value, $n + 1$. For more efficient coding, the refinement JPEG-LS coder should be modified so it reconstructs the near parameter for each pixel and uses this information in coding the difference pixel. Compared with other results,[14] our versions have the advantage of a better control of the (maximum) error of the lossy coding and if using visual quantization, a finer granularity. It is possible to use near-lossless factor up to 5 without any significant visual losses.

For example the simple scalable coding mentioned based on selecting k LSB planes[14] has the following drawbacks:

1. The intervals steps increase exponentially as $2^k$, i.e. 2, 4, 8, 16, ... .
2. As can be seen from the result,[14] the convex hull of the rate-distortion (measured by PSNR) of the baselayer has lower quality than the convex hull of JPEG-LS measured at different near values.
3. Also measured by the inf-norm truncation it is inferior to using JPEG-LS in near-lossless mode. (Points 2 and 3 are in part due to reconstructing to integer points.) Due to these relative advantages of JPEG-LS in near-lossless mode, it is interesting to explore the possibilities of using JPEG-LS near-lossless directly as the base-layer, but still have a simple coding and processing of enhancement layers for a scalable codec.

## 4. COMPARISON METHODOLOGY

To decrease the level of power consumption and complexity cost it is suggested here to use video encoders with relatively small level of operational complexity and very small memory utilization.

### 4.1. Intra-Coding and Slicing

To decrease complexity costs at the encoder side intra-coding mode only is used, i.e. the frames are compressed independently and are not stored after compression. Therefore we can not apply algorithms that exploit temporal redundancy such as motion estimation/compensation (ME/MC) or differential coding.

An additional way to decrease memory consumption for some algorithms like JPEG2000 is to split an input frame into smaller pieces and to process each piece independently and sequentially. More specifically, a frame is partitioned into (one or) several disjoint rectangular regions called slices. Small slices can

significantly decrease the level of memory consumption. The slice size in our tests are 1280x16 (less than 100 KByte if a pixel depth is 24 bits) and for hardware implementation a low-cost internal chip-based memory can be used. Slicing can be viewed from an implementation point of view: we have one slice in memory at a time. This will restrict wavelet based coders, whereas JPEG-LS only requires 2-lines in memory (nevertheless, coded as one frame). Meanwhile, slicing can also be viewed from the point of the coding format, which may influence performance. For JPEG-LS we can code the whole image as one entity (still not requiring more than two lines in memory) or we can slice it. The later approach for slicing will come at a price in coding performance for adaptive coders as JPEG-LS. However, slicing also allows us to enhance the use of unequal error protection (UEP)[19] to the transmission of the video data.

### 4.2. Scalability

Data transmission with two sub-streams is one of the most commonly used scalability schemes satisfying most of the use cases. It allows for end-users to receive pictures even in the case of unexpected data loss and at the same time the overhead costs for (unequal) error protection for two streams instead of one are quite acceptable. So, the main idea of scalable coding is that coder forms the bitstream from several layers. All three competitors support scalable and progressive coding i.e. in case of partial data loss during transmission only a part of the encoded data could be received, extracted and decoded from a global scalable bit-stream and an image with smaller quality level can be successfully reconstructed at the receiver. Scalability is one of the most prominent features of JPEG2000 that leads to its ability of extracting different resolutions, fidelities, components or spatial locations from a single compressed bitstream. The scalable extension (SVC) of the H.264/AVC Standard, is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems. Current version of JPEG-LS standard (excluding the one proposed in this paper) does not support scalability.

### 4.3. Complexity Costs and Features

It is clear that implementation complexity and memory consumption should be estimated not for the whole coding standard but for a selected core number of functionalities required by wireless applications. The standardized codecs provide a rich set of instruments for putting certain restrictions on the encoding parameters such that some kind of complexity scalability can be achieved. Anyway different investigations[2] demonstrate that JPEG2000 is more complex than H.264/SVC due to integral arithmetic coding and multiple bitwise operations. It was mentioned[20,21] that the Tier-1 block of JPEG2000 that includes those operations consumes more than 50% of total computation power. On the other hand JPEG-LS was originally designed as a low-complexity coding system that has extremely low complexity level in comparison with H.264/SVC and JPEG2000. Error concealment in JPEG2000 can be ensured by different means including markers, regular termination of the arithmetic coder, error resilient termination and segment symbols, possibility to move the sensitive packet head information to the bit stream header and etc .[4] H.264/SVC also includes a rich set of tools particularly designed for that purpose.[5] JPEG-LS unfortunately does not offer a proper error resilience support, but as described may easily be extended..

## 5. RESULTS

In our experiments HD video sequences of resolutions 1280x720, 30fps are tested. They have a 24 bits/pixel depth. The raw bitrate is 0.663 Gbps. We have used RGB 4:4:4 color space format for JPEG2000 and JPEG-LS and YUV 4:4:4 for H.264/SVC. For more detailed description of the codecs settings please refer to Subsection 2.4. "Kungfu" (427 frames) and "Breeze" (461 frames) are natural image sequences, with fast and slow motion, respectively. "Desktop" (1880 frames) is a computer desktop snaphot, mostly consisting of computer graphics. (For the sake of simplicity and ease of reporting on the figures lossless is depicted at 70dB is for JPEG-LS and JPEG2000). The following graphs (Figs.2-4) present the comparison of results. It is necessary to note that one point for JPEG-LS (between n=0 and n=1) is obtained as a compression of one half of the frames with n=0 and the other - with n=1.

Obviously, JPEG-LS was not constructed for the low rates, in contrast to H.264. JPEG-LS has the best relative performance for computer graphics, that is why it shows outstanding results on the "Desktop"

sequence (Fig.2). Table 1 shows compression efficiency comparison for Standard JPEG-LS and the proposed scalable solution in Gbps.

**Table 1.** Compression efficiency comparison (bit rate, Gbps)

| Codec type | desktop | kungfu | breeze |
|---|---|---|---|
| JPEG-LS | 0.128 | 0.229 | 0.262 |
| Scalable JPEG-LS | 0.132 | 0.247 | 0.276 |

## 6. CONCLUSION

This paper proposes simple and low-complexity solution for two-step scalable JPEG-LS. Comparison with other widely used scalable solutions for video compression like H.264/SVC and JPEG2000 show that the proposed idea can compete with standard solutions for specific video content at high rates.

## 7. ACKNOWLEDGMENT

This research work was supported in part by a special grant of Intel CTG Research Council. We are especially grateful to Dr. Andrew Belogolovy and Dr. Andrey Turlikov for plentiful discussions and helpful comments.

## REFERENCES

1. ISO/IEC JTC 1/SC 29/WG1 FCD 14495-public draft, "JPEG-LS: lossless and near-lossless coding of continuous tone still images," July 16th, 1997
2. D. Santa Cruz et al., "An analytical study of JPEG 2000 functionalities," *Proceedings of SPIE*, Vol. 4115, 2000.
3. S. D. Rane, G. Sapiro, "Evaluation of JPEG-LS, the New Lossless and Controlled-Lossy Still Image Compression Standard, for Compression of High-Resolution Elevation Data," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 10, October 2001.
4. ITU-T and ISO/IEC JTC 1, "JPEG 2000 Image Coding System: Core Coding System, ITU-T Recommendation T.800 and ISO/IEC 15444-1," *JPEG 2000 Part 1*, 2000.
5. "Introduction to SVC Extension of Advanced Video Coding, ISO/IEC JTC1/SC29/WG11" *International Organization for Standardization, Coding of Moving Pictures and Audio,* Pozna , Poland, July 2005. http://www.chiariglione.org/mpeg/technologies/mp04-svc/svc/.
6. D. S. Taubman, "Directionality and Scalability in Image and Video Compression," *PhD dissertation,* 1994.
7. K. N. Ngan, C. W. Yap, K. T. Tan "Video coding for wireless communication systems," Marcel Dekker, 2001.
8. A. Belogolovy, E. Belyaev, A. Sergeev, A. Turlikov, "Video Compression for Wireless Transmission: Reducing the Power Consumption of the WPAN Hi-speed Systems," *New2AN*, 2009.
9. A. H. Sadka "Compressed video communications," John Wiley and Sons, 2002
10. ISO/IEC 10918-1 ITU-T Recommendation T.81) http://www.w3.org/Graphics/JPEG/itu-t81.pdf
11. T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003.
12. M.J. Weinberger, G. Seroussi, G.Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *Image Processing, IEEE Transactions*, Volume 9, Issue 8, Aug 2000 Page(s):1309 - 1324.
13. E.Belyaev, V. Grinko, A. Ukhanova, "Power Saving Control for the Mobile Receivers in the DVB-H based on the Scalable Extension of H.264/AVC Standard," *Wireless Telecommunications Symposium*, 2009
14. R.J.Vleuten, S. Egner, "Lossless and Fine-Granularity Scalable Near-Lossless Color Image Compression," *Data Compression Conference,* 2002, DCC Proceedings.
15. D.G. Korn, K.P. Vo, B. Krishnamurthy, "Vdelta: Differencing and Compression, Practical Reusable Unix Software" John Wiley Sons, 1995.

16. D. S. Taubman, "Directionality and Scalability in Image and Video Compression," *PhD dissertation,* 1994.
17. "DPCM - Differential Pulse Code Modulation," http://einstein.informatik.uni-oldenburg.de/rechnernetze/dpcm.htm.
18. A. Krivoulets, "Design of efficient algorithms for image compression with application to medical images," *Ph.D. thesis at IT University of Copenhagen,* ISBN: 87-7949-060-3, 2004.
19. R.M. Joohee Kim Mersereau, Y. Altunbasak, "Error-Resilient Image and Video Transmission Over the Internet Using Unequal Error Protection," Transactions on Image Processing, Vol. 12, No. 2, pp. 121 - 131, Ferbuary 2003.
20. Chen, Liang-Gee, *United States Application 20060008162,* http://www.freepatentsonline.com/20060008162.html.
21. G. Pastuszak, "A high-performance architecture for EBCOT in the JPEG 2000 encoder," *Signal Processing Systems Design and Implementation,* IEEE Workshop on Volume , Issue , 2-4 Nov. 2005 pp. 693 698.
22. B. Martins and S. Forchhammer, "Lossless Compression of Video using Motion Compensation,"
23. S. Andriani, G. Calvagno, M. Durigon, R. Rinaldo, M. Knee, P. Walland, M. Koppetz, T. Erseghe, G.A. Mian, "Comparison of lossy to lossless compression techniques for digital cinema," *International Conference on Image Processing*, Singapore, 2004.
24. D. Brunello, G. Calvagno, G.A. Mian, and R. Rinaldo, "Lossless Compression of Video Using Temporal Information," *IEEE Transactions on image Processing*, Vol. 12, No. 2, February 2003.

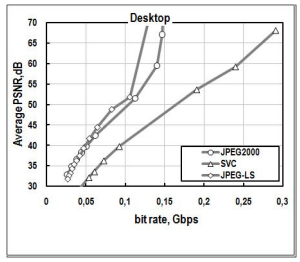



Figure 2. RD comparison for "Desktop" sequence

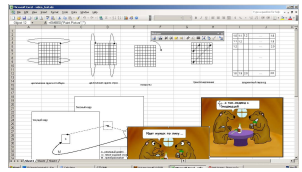


Figure 5. Example of "Desktop" sequence

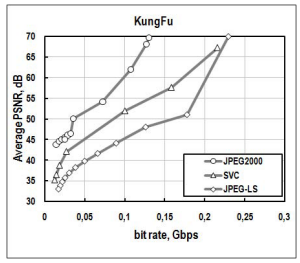



Figure 3. RD comparison for "Kungfu" sequence

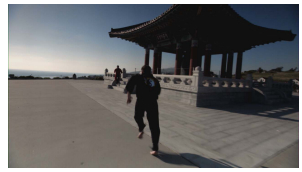


Figure 6. Example of "Kungfu" sequence

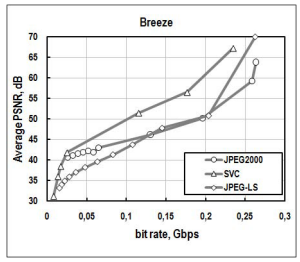



Figure 4. RD comparison for "Breeze" sequence

Figure 7. Example of "Breeze" sequence

# Low-complexity JPEG-based progressive video codec for wireless video transmission

Anna Ukhanova, Søren Forchhammer
DTU Fotonik, Technical University of Denmark

***Abstract*—This paper discusses the question of video codec enhancement for wireless video transmission of high definition video data taking into account constraints on memory and complexity. Starting from parameter adjustment for JPEG2000 compression algorithm used for wireless transmission and achieving the best possible results by tuning settings, this work proceeds to develop a low-complexity progressive codec based on JPEG, which is compared to the tuned JPEG2000. Comparison to H.264/SVC for this codec is also given. As the results show, our simple solution on low rates can compete with JPEG2000 and H.264/SVC for specific video content.**

## I. Introduction

The parallel evolution of wireless communication systems and video data compression schemes has reached the level where it seems to be realistic to transmit high definition video sequences with variable quality adapting to fluctuating wireless channels. The variety of multimedia applications forms individual requirements for the video compression algorithms. Moreover, the organization of the encoded bit stream depending on the transmission protocols is also important for the quality of the reconstructed video sequence. Finally, the commercial use of these multimedia applications means that there could be strong memory and computational complexity constraints. One of the most investigated and popular cases is a real-time high quality video transmission in wireless personal area networks (WPANs [1]) in houses and offices. Applications and technical specifications aim to deliver the high-definition video (HD video) within the office or entertainment cluster wirelessly to the huge variety of devices: digital still cameras, HDTV, set top boxes, game consoles, camcorders, high resolution printers/scanners, players, projectors and etc. Applications include cable replacement, remote connection to HD displays, multimedia exchange within offices, and video transmission from informational kiosk. For example, UWB [2] could connect a wall-mounted plasma display to an STB or DVD player without any cables (see Fig. 1). Set-top-box resources are enough to transcode or recompress video to fit narrow client channels. One of the key points of wireless transmission is a fluctuative nature of the channel. Therefore, progressive video coding should be used to provide continuous playback on the receiver side. Hovewer, current solutions [3] can not be considered to be simple in terms of computational/memory complexity (e.g. [4]). Industrial use and commercial implementation of these algorithms has already put some restrictions on the cost of the end-user devices for wireless video transmission meaning it is necessary to decrease complexity of the algorithms and, therefore, decrease the overall price of the device. For initial decrease of the complexity costs we focus on frame by frame coding of video as in M-JPEG [5], M-J2k and Intra coding only in H.264/SVC. But as an alternative to the existing algorithms we decided to develop low-complexity solution for simple progressive encoding algorithm based on DCT. A low latency constraint influences the development of the algorithm and, as stated above, it was decided to use intra-coding only as restrictions like fixed bitrate combined with the use of motion compensation would most likely increase latency. Also frame by frame coding is more robust compared to motion compensation. Progressiveness also allows to apply unequal error protection [6]. We adjust our algorithm for HD video compression so that it does not perform worse than JPEG2000 [7] for particular video content (e.g. HD videogaming as a part of home network environement).

The paper is organized as follows. Firstly, we suggest some simple codec adjustments to the specific use cases for the generally recognized leader in the area of image compression: JPEG2000 (the full-featured international standard for still images coding). The goal is to achieve good results under constrained complexity. Secondly, the specific video content types are discussed and different ways for compression of these types are given. Then the solution for low-complexity tiny codec with the feature of progressiveness is presented. RD-performance of the progressive tiny codec is thereafter given along with

9781–4244–7286–4/10/$26.00 ©2010 IEEE

A.Ukhanova, S. Forchhammer, “Low-complexity JPEG-based progressive video codec for wireless video transmission” , *International Conference on Ultra Modern Telecommunications (ICUMT)*, Moscow, Russia, 2010.

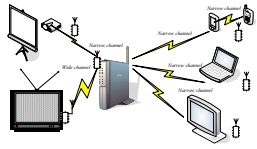



Fig. 1. Wireless video transmission use cases

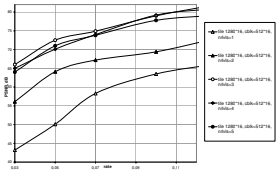


Fig. 2. Encoding parameters adjustment for JPEG2000 for "video" type of the sequence

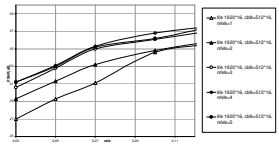


Fig. 3. Encoding parameters adjustment for JPEG2000 for "synthetic" images

results for JPEG2000 and H.264/SVC [8]. Finally, the computational and memory complexity of the proposed codec is calculated and compared to JPEG2000.

## II. Compression algorithms parameters

Existing algorithms have different characteristics because the goal and period of their creation was also different. For instance, JPEG [5] has low complexity and good compression ratio whereas JPEG2000 has very good compression ratio at the expense of a rather high complexity. We focus on the development of algorithms for video compression for specific constrained use cases, where existing image and video compression algorithms, adapted to general conditions, fail to give the best performance and could be improved.

### *A. Parameter adjustment*

The first approach is based on parameter adjustment without any change in the encoding algorithms and without adding complementary blocks to it. We try to tune parameters of JPEG2000 [9] algorithm based on discrete wavelet transform (DWT) to find the best settings for compression of different types of the video sequences in order to further compare "optimized" JPEG2000 with the proposed codec.

### *B. Dividing pictures into tiles for memory consumption*

If we have strong memory constraints at the encoder/decoder side, for JPEG2000 algorithm we could use tiling to provide image partitioning into rectangular and non-overlapping tiles of size, e.g. $1280 \times 8$ or $1920 \times 8$. The image quality at a given bit-rate is reduced in this case. This happens because, firstly, the coding algorithm (and, in particular, the DWT) is applied not to the whole image and, secondly, other coding parameters (e.g. codeblock size, number of resolution levels) are not matched with the tile size, but they remain at their default values.

### *C. Optimal parameters for tiling*

By coordinating parameters with each other we can achieve a better quality (as shown on Fig. 2, for the test videosequence "Kungfu", 427 frames, $1280 \times 720$). As seen from the results, the best mode with tiling option is the biggest supported codeblock size $512 \times 16$ for tilesize $1280 \times 16$ and the number of DWT levels equal 3 (nrvls=4 for Jasper [10] codec) for "video" type of the sequence. Another mode with 2 DWT levels should be chosen for compressing computer (synthetic) images (see Fig. 3), which are common for video games and computer graphics. The selected test set in this case is a sequence "Synthetic", ($1920 \times 1080$, 15 frames).

### III. Detection of "text"/"video"

The second approach consists of using some extra blocks in the encoding algorithms, which help to improve compression efficiency. A good example of these blocks is a "text"/"video" detector. Some images like computer desktop screenshots actually consist of two types of the images - photorealistic images or so called "video" type of the images (or tiles, when we divide a picture into non-overlapping blocks), and computer images with text (mostly, uniform background of one colour with letters written in the other colour) - the so called "text" type. The detector helps to divide images into non-overlapping blocks of text and video types. A simple approach is described below. The developed algorithm detects text areas in tiles of the input image. The aim of the detector is to separate input tile into sets of "text" or "video" segments. Each segment is then compressed with the optimal parameters (number of DWT levels for JPEG2000 and the biggest supported codeblock size). These parameters could be applied to the detected areas, which are the parts of the tiles and as each tile is anyway compressed independently, we can divide the tile into sub-areas of "text" and "videos". Icons, toolbars and other elements of screen captures should be detected as text. The tile is divided into blocks of size N×N. Further the number of different colors is counted in each block. Thus, a virtual color map is produced according to the initial image (its size - $image\_height/block\_height \times image\_width/block\_width$). Each cell of this map contains the information about the number of colors of an appropriate block. The next step is a classification of the blocks. The number of colors (NC) in each block is compared with a threshold ($T$, e.g. $T$=3). And if NC is less then $T$, the block is classified as "text" block, otherwise as a "video" block. Then the algorithm works with each block and its three neighbors. If NC of current block or at least one of its neighbors is less then $T$, all blocks (current block and its three neighbors) are classified as "text" blocks. The result of the algorithm work one can see on Fig. 4.

By differentiating two image types, we can achieve a big gain. One can see that the compression efficiency degrades a lot when we use optimal video coding parameters (number of DWT levels = 2) for "text" (no DWT for "text" images) and vice versa (results are estimated for the videosequences "Desktop", 1880 frames, and "Kungfu", 427 frames, both $1280 \times 720$ ). It can be noticed that quality decreases much more when "text" ("Desktop") is compressed as "video" - see Fig. 5 - ( 20 dB for rate=0.13) than for inverse situation when "video" ("Kungfu") is compressed as "text" - see Fig. 6 - ( 4 dB for rate=0.13).

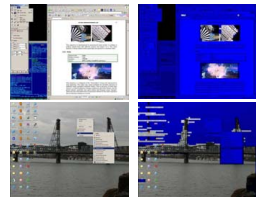


Fig. 4. "Text" and "video" detection. Original image (left) and detected (right) (N=16, T=3)

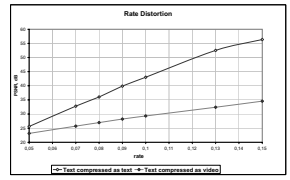



Fig. 5. "Text" images compressed with different parameters - "Desktop" videosequence

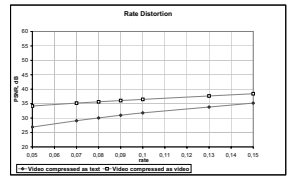



Fig. 6. "Video" images compressed with different parameters - "Kungfu" videosequence

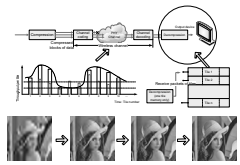


Fig. 7. Scheme of the progressive transmission

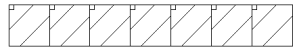


Fig. 8. Partitioning into slices

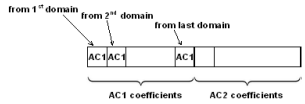



Fig. 9. Coefficients ordering in slices

## IV. Progressiveness: proposed algorithm

For increased performance, we can also change some blocks inside the encoder. For example, one of the important characteristics of compression algorithms is progressiveness. The JPEG and JPEG2000 standards also specify progressive coding but at the cost of increase in complexity. Here, we propose a simpler solution for progressive image compression based on JPEG.

### *A. What is actually progressiveness?*

When we have to transmit data - video sequences - in real-time, it is very important that we are able to decode (and, therefore, display) the image (rows of image) in each set of time. If Standard (non-progressive) JPEG [5] is used for video encoding and real-time transmission, channel errors of even one packet can cause fatal problems for a decoder, and a corresponding part of the image (e.g. domain) will not be represented. Therefore, a progressive version of image compression algorithm is desired for that case. Progressiveness allows to see a rough approximation of the input image after the first stream has been decoded, and the image reproduction quality is then gradually improved as more scans are decoded (see Fig. 7)

### *B. Proposed Progressive Tiny (pTiny) algorithm*

For HDTV images it is not efficient to encode the domains (i.e. image blocks) separately. It is better to scan the domain coefficients of the image rows in a particular order (Fig. 8, "slice scanning").

We can gather the coefficients on the same positions as subbands. This grouping into subbands will also allow us to provide progressiveness [11]. In other words, we scan all first coefficients of all domains in one row, then the second ones, then the third etc., for each slice (Fig. 9). Thereby the coefficients of approximately equal values will be collected together (that is most valuable for HDTV images).

We have developed the compression algorithm called Progressive Tiny which is less complex than the usual Progressive JPEG [12] and less complex than JPEG2000 both for software and hardware implementation. For the case of memory restrictions it gives results which in some cases may be comparable to those of JPEG2000 with adjusted parameters for tiling and quite close to the results to the H.264/SVC (Fig. 12–14). More discussions are given later.

To achieve a low memory consumption, the encoder and decoder for our pTiny algorithm has input storage only for one row of domains for Y, U and V components. For the Y component we have $(image\_height)/(domain\_height)$ rows in the image. For U and V components this number reduced to half because of downsampling. We form packets simultaneously for Y, U and V components, and should somehow figure out the problem caused by downsampling. There are two ways of solving this problem: we can either have a bigger input storage, so that we can store two rows of Y, one of U and one of V at the same time (YYUV), or change the procedure of forming the domain (YUV).

To reduce the memory consumption, we decided to have specific domains forming procedure for U and V components of image. This procedure does not affect the visual quality significantly (see Fig. 10), but allows to

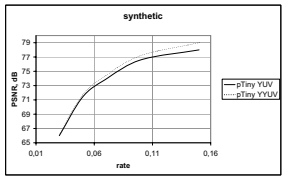



Fig. 10. Comparison of YYUV and YUV schemes of domain forming

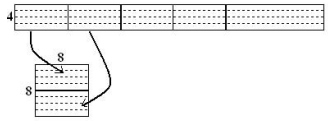



Fig. 11. Forming UV-domains

reduce the memory consumption by a factor of 2. We have only 4 rows of pixels of the U and V components, and we have to construct a domain of $8 \times 8$ pixels. So, we construct this domain in the following manner (see Fig. 11)

In the pTiny algorithm the rows of the image are firstly divided into domains ($8 \times 8$). Then these domains are DCT-transformed and quantized by a special quantization matrix that could be implemented in hardware by using shift operations only.

After these transformations we extract DC coefficients from each domain in row and create first slice. Then the difference between DC coefficients will be calculated and this slice will be encoded with the standard Huffman tables for DC.

Then we form slices of AC coefficients according to the established division into slices using the slice scanning algorithm. It is clear that we can combine two scanning algorithm (i.e. ZigZag scanning and slice scanning) in one without any losses.

We use the same slice scanning algorithm for all slices and finally have a new array of AC coefficients, that actually has the same length (with DC coefficients), as the whole row of domains.

The advantage of such a rearrangement of coefficients is that we group zero coefficients from all domains in one row together. It is expected that there are lots of zeros in the end of the domains, so that some of the last slices completely consist of zeros. And for images of big width (HDTV size) they can be effectively compressed. The same approach is presented by the traditional ZigZag scanning in Standard Sequential JPEG algorithm.

After this scanning we can apply a modified algorithm for Run Length Encoding, which works as follows:

1) For traditional pairs of Run and Size, when Run is smaller than 16, we use traditional pairs (Run, Size)(Amplitude)
2) When Run is 16 or more, we use ZeroBlock coding $(ZeroBlock)(ZeroRun)$, where $ZeroBlock$ represents number of bits used to encode $ZeroRun$, $ZeroRun$ is a zero-run value. $ZeroBlock$ is represented in pairs as following (number of bits, 0)
3) When the last run of zeros includes the last AC coefficient in the slice, we use a special symbol EOS (End Of Slice), that has value (0,0)

Traditional Huffman tables for pairs (Run, Size) stated in JPEG Standard do not include ZeroBlock values. Moreover, they are general for all images. It is obvious, that for fixed parameters (image size, quality) we can create more optimal Huffman code. So we use our own tables for entropy coding of (Run, Size). One table is for Y component of the image, the other one is for U and V components together. These tables were created by applying Huffman algorithm to the statistic for pairs (Run, Size) that was collected over big amount of images (mostly computer graphics of HDTV size, and also some natural HDTV images ).

## V. Performance comparison

Figures 12–14 show the compression performance of three fully progressive algorithms, namely JPEG2000, H.264/SVC and pTiny. JPEG2000 was applied with the tiling option with tile height = 16, and adjusted parameters for this tile size and video sequence characteristics. H.264/SVC was used in Intra-mode to provide a fair comparison with M-JPEG based pTiny. One can see that JPEG2000 and H.264/SVC perform better on the "video" sequence "Kungfu" as originally these algorithms were created for photorealistic image/video compression. On the "text" sequence "Desktop" JPEG2000 has a significant gain due to the parameters adjustment for this type of the image. However, pTiny achieves almost the same compression results within the ranges of rates 0.03-0.05. pTiny shows good results on "synthetic" video type as

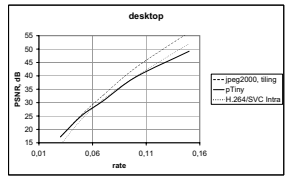



Fig. 12. Compression efficiency comparison of the Progressive Tiny (pTiny) algorithm on "text" sequence

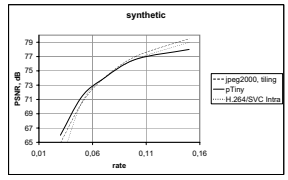



Fig. 13. Compression efficiency comparison of the Progressive Tiny (pTiny) algorithm on "synthetic" sequence

both slice scanning algorithm and proposed Huffman tables were tuned for this type of the images.

## VI. Complexity issues

In the previous section performance capability has been adressed. In this part complexity of the proposed algorithm is analyzed and compared with the corresponding issues for JPEG2000. The fundamental building blocks of a typical JPEG2000 encoder can be joined into two major parts:

- pre-processing, DWT, quantization;
- Embedded Block Coding with Optimized Truncation (EBCOT): arithmetic coding (Tier-1 coding), and bitstream organization (Tier-2 coding). Tier-1 is an embedded block coder, which utilizes context analysis and context-based arithmetic coding to encode each code-block into an independent embedded bit-stream. Tier-2 is a post-compression rate-distortion optimization algorithm. EBCOT Tier-1 is the most complex part of JPEG 2000, which consumes more than 50% of the total computation power [13].

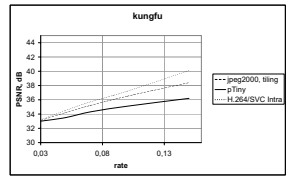



Fig. 14. Compression efficiency comparison of the Progressive Tiny (pTiny) algorithm on "video" sequence

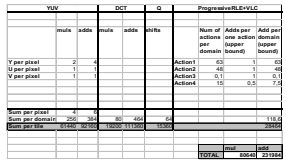
Our research shows that the JPEG2000 coding system is full-featured but much more complex than JPEG-based solution. Anyway there are 2 main ways for complexity reduction of JPEG2000:

- decrease the number of coding passes of context-based arithmetic coder if rate control [14] is applied (the most computational-intensive part of JPEG 2000). Algorithm should be developed to stop context-based arithmetic coder;
- find optimal DWT filters for usage with tiling and reduce the size of vertical DWT filter.

We tried to estimate operational complexity in terms of multiplications (muls), additions (adds) and shifts and memory consumption of pTiny codec. The calculated complexity level (see Fig. 15) per tile for HDTV ($1920 \times 1080$) is an "upper bound" and can be significantly improved during final design process. Firstly, complexity of entropy coding of the pTiny algorithm is estimated in different Actions (Action1 = Comparison with 0, Action2 = Counter increasing, Action3 = Counter modifying, Action4 = Amplitude calculating (only for negative sign)), and then pseudo-tranformed into additions.

### *A. Color space tranform*

The input video sequence in RGB color space is converted to YUV (luminance and 2 chrominance components). In the pTiny codec YUV 4:1:1 (4:2:0) reduction is applied. For every 4 points of the U and V color components an average value is stored. So



| | YUV | | DCT | | Q | ProgressiveRLE+VLC | | |
|---|---|---|---|---|---|---|---|---|
| | muls | adds | muls | adds | shifts | Num of actions per domain | Adds per one action (upper bound) | Add per domain (upper bound) |
| Y per pixel | 2 | 4 | | | | Action1 | 63 | 1 | 63 |
| U per pixel | 1 | 1 | | | | Action2 | 46 | 1 | 46 |
| V per pixel | 1 | 1 | | | | Action3 | 0,1 | 1 | 5,1 |
| | | | | | | Action4 | 15 | 0,5 | 7,5 |
| Sum per pixel | 4 | 6 | | | | | | | |
| Sum per domain | 256 | 384 | 80 | 464 | 64 | | | | 116,6 |
| Sum per tile | 81440 | 32160 | 19200 | 111360 | 15360 | | | | 29456 |
| | | | | | | | mul | | add |
| | | | | | | TOTAL: | 89640 | | 231984 |

Fig. 15. Calculation of pTiny complexity

the overall complexity of this stage is the following: 9 bytes to store color space transform coefficients (memory consumptions) using 9 multiplications, 6 additions and 2 shifts per RGB sample. For $8 \times 8$ domain the maximum complexity is $64\times$(9 muls + 6 adds + 2 shifts) = 576 muls + 384 adds + 128 shifts.

*B. Discrete cosine transform (DCT)*

DCT is a very important aspect of a JPEG-based scheme. Decorrelation of the coefficients is very important for compression, because each coefficient can then be treated independently without loss of compression efficiency. Another important aspect of the DCT is the ability to quantize the DCT coefficients using visually-weighted quantization values. The most straightforward way to implement DCT is to follow the theoretical equations. When we do this, we get an upper limit of 64 multiplications and 56 additions for each 1-D 8-point DCT. (Note that the cosine terms can be combined with the constants before the computation because the cosines become discrete numbers at each position.) Therefore, a full $8 \times 8$ DCT done in this way in separable 1-D format - eight rows and then eight columns - would require 1024 multiplications and 896 additions plus additional operations to quantize the coefficients. If the DCT really required this many operations to compute, JPEG would have chosen a different algorithm. In fact, the most efficient algorithm for the $8 \times 8$ DCT, a true 2-D method developed by Feig [15] requires only 54 multiplications, 464 additions, and 6 arithmetic shifts to produce a form suitable for quantization. But it requires float quantizers after FDCT to scale output. The version of the DCT by Arai, Agui, and Nakajima [15] has 13 multiplications and 29 additions, which is competitive with the best that has been achieved by other techniques. What makes it special, however, are the properties it has for a scaled DCT implementation. Although 13 multiplications are required, eight of the multiplies are to scale the final output to the correct range as in Feig scheme. But for this algorithm scaling can be done by shift operations. If the output is to be quantized - as in the case of JPEG - the outputs can be left in this scaled form and the scaling factors can be absorbed into the divisions needed to quantize the outputs. Only 5 multiplications are actually needed before the quantization, making this the most efficient 1-D quantized DCT known. Because the 2-D DCT is separable, the summations can be done as eight 1-D DCTs on all rows, and then eight 1-D DCTs on the eight columns formed by the coefficients for the rows. Note, however, that if we use the scaled 1-D DCT we can delay the row output scaling until after the column 1-D DCTs are complete. Consequently, we can embed the output scaling for both the row DCTs and the column DCTs in the final quantization step. To get to the final scaled DCT output we therefore need 16 1-D scaled DCTs, and therefore a total of $16 \times 5$ = 80 multiplications and $16 \times 29$ = 464 additions. This is the best approach known for a separable 2-D scaled DCT [15]

*C. Quantization*

One of the simpliest ways to decrease JPEG complexity is to use only one predefined quantizing table for the target quality level. Moreover only shift operations are used in the proposed scheme instead of the divisions in the standard quantizing tables (the proposed quantization table of pTiny codec consists of powers of 2).

*D. Entropy Coding*

It is useful to consider entropy coding as a 2-step process. The first step converts the zigzag sequence of quantized coefficients into an intermediate sequence of symbols. The second step converts the symbols to a data stream in which the symbols no longer have externally identifiable boundaries. The form and definition of the intermediate symbols is dependent on both the DCT-based mode of operation and the entropy coding method. There are several ways of coding these sequences, using Run Length coding and Huffman coding. Figure 15 shows average complexity values per block for the operations listed above. For example if average Action-3 complexity equals 0.1 it means that there is only one Action-3 per 10 blocks in average.

Assume that JPEG2000 applies a color space transform, which has the same complexity. For the DWT algorithm with filter 5-3, it is necessary to have 8 additions per pixel, $8 \times 64$ for domain, that is similar for pTiny.

Even if we consider that JPEG has the same complexity of the quantization, taking into account that entropy coding part takes 50 % of the complexity (furthermore reading bitplanes is hard for hardware implementation), it means that JPEG2000 is more complex than proposed solution. Table 1 also shows comparison of the memory consumption.

TABLE I
MEMORY SIZE COMPARISON

| Algorithm | J2k $16 \times 1920$ | PTiny $8 \times 1920$ |
|---|---|---|
| Memory (kB) | 150 | 50 |

## VII. CONCLUSION

To meet the requirements of real-time video applications and to decrease the implementation complexity, the proposed pTiny codec was tuned for HD computer graphics only. The work focused on the entropy coding scheme. An entropy coding method with reduced complexity was proposed. The compression algorithm pTiny was compared with other progressive codecs as JPEG2000 and H.264/SVC. JPEG2000 was also adjusted to specific conditions caused by computational complexity and memory restrictions. Thus, it was possible to achieve a big gain in compression efficiency by setting different parameters for the ”text” and ” video” parts of the images compared to default parameters. The proposed pTiny algorithm allows to provide good compression performance mostly on HD computer graphics (”synthetic” type) that could be explained by using this kind of images for gathering statistics for tuning the entropy coding and adjusting slice scanning algorithm exactly on this type of images. Progressive decoding could be applied to the code stream of this algorithm, as is also the case for H.264/SVC and JPEG2000. It allows to use unequal error protection for different slices of the output data, and also to decode the video data if only some part of the bitstream was received. The analysis and experimental results show that the proposed solution significantly decrease codec complexity (70 % memory consumption) while compression/quality ratio only changes slightly.

## ACKNOWLEDGMENT

The authors would like to thank SUAI Video Lab for supporting, especially Andrey Turlikov and Anton Sergeev for their contribution to this work.

## REFERENCES

[1] 802.15 IEEE standards: http://www.ieee802.org.
[2] "IEEE 802.15 WPAN Millimeter Wave Alternative PHY Task Group 3c (TG3c)," Contributions and documents, 2009.
[3] H. Schwarz, D. Marpe, and T. Wiegand "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9, September 2007.
[4] JSVM 9.15 software package, CVS server for the JSVM software: http://iphome.hhi.de/
[5] ISO/IEC 10918-1 ITU-T Recommendation T.81 http://www.w3.org/Graphics/JPEG/itu-t81.pdf
[6] Joohee Kim Mersereau, R.M. Altunbasak, Y. "Error-Resilient Image and Video Transmission Over the Internet Using Unequal Error Protection," IEEE Transactions on Image Processing, Vol. 12, No. 2, pp. 121 - 131, Ferbuary 2003.
[7] ITU-T and ISO/IEC JTC 1, "JPEG 2000 Image Coding System: Core Coding System, ITU-T Recommendation T.800 and ISO/IEC 15444-1," JPEG 2000 Part 1, 2000.
[8] H. Schwarz, M.Wien, "The Scalable Video Coding Extension of the H.264/AVC Standard" *IEEE Signal Processing Magazine*, vol. 25, Is. 2, pp. 135-141, 2008.
[9] D. S. Cruz et al., "An analytical study of JPEG 2000 functionalities," ISO/IEC JTC1/SC29/WG1 N1816 Coding of Still Pictures, July 2000.
[10] JPEG2000 Software, http://www.ece.uvic.ca/ mdadams/jasper/
[11] A. Cheng, F. Shang, "Priority-Driven Coding of Progressive JPEG Images for Transmission in Real-Time Applications," 11th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'05), 2005.
[12] J. In, S. Shirani, F. Kossentini, "On RD optimized progressive image coding using JPEG," IEEE Transactions on Image Processing, vol. 8, issue 11, 1999
[13] G. Pastuszak, "A high-performance architecture for EBCOT in the JPEG 2000 encoder," Signal Processing Systems Design and Implementation, 2005.
[14] D. Taubman and M. W. Marcellin, "JPEG2000: Image Compression Fundamentals, Practice and Standards," Massachusetts: Kluwer Academic Publishers, 2002.
[15] W. B. Pennebaker, J. L. Mitchel, "JPEG still image data compression standard," NY, 1993

# Scalable-to-Lossless Transform Domain Distributed Video Coding

Xin Huang #, Anna Ukhanova #, Anton Veselov *, Søren Forchhammer #, Marat Gilmutdinov *

# *DTU Fotonik, Technical University of Denmark, Building 343, Lyngby 2800, Denmark*

* *Saint-Petersburg University of Aerospace Instrumentation. 67, Bolshaya Morskaya str., Saint-Petersburg, Russia*

***Abstract*—Distributed video coding (DVC) is a novel approach providing new features as low complexity encoding by mainly exploiting the source statistics at the decoder based on the availability of decoder side information. In this paper, scalable-to-lossless DVC is presented based on extending a lossy Transform Domain Wyner-Ziv (TDWZ) distributed video codec with feedback. The lossless coding is obtained by using a reversible integer DCT. Experimental results show that the performance of the proposed scalable-to-lossless TDWZ video codec can outperform alternatives based on the JPEG 2000 standard. The TDWZ codec provides frame by frame encoding. Comparing the lossless coding efficiency, the proposed scalable-to-lossless TDWZ video codec can save up to $5\%$-$13\%$ bits compared to JPEG LS and H.264 Intra frame lossless coding and do so as a scalable-to-lossless coding .**

## I. Introduction

Distributed Video Coding (DVC) [1] is a new video coding paradigm, which mainly exploits the source statistics at the decoder instead of at the encoder as in motion-compensated video encoding. Thereby computational power requirements are shifted from encoder to decoder. According to the Slepian-Wolf theorem [2], it is possible to achieve the same rate by independently encoding but jointly decoding two statistically dependent signals as for typical joint encoding and decoding (with a vanishing error probability). The Wyner-Ziv theorem [3] extends the Slepian-Wolf theorem to the lossy case. This work laid the basis for distributed source coding and it forms the key theoretical basis for DVC. The source is lossy coded based on the knowledge that a correlated source is available at the decoder, which utilizes this so-called side information. There are various approaches to DVC, e.g. PRISM [4], Pixel Domain Wyner-Ziv (PDWZ), and feedback channel based Transform Domain Wyner-Ziv (TDWZ) [5] have been applied for lossy coding. However, the lossless distributed source, image and video coding has also been devised, [6]-[9]. One application considered is for hyperspectral images [7]-[9], but lossless distributed source coding may also find relevant applications in other scientific and medical applications. A good example of the need for lossless coding could be scanned image sequences for 3D reconstructions. There could be several approaches for providing lossless DVC. For example, a wavelet based DSC approach for lossy-to-lossless compression of hyperspectral images is proposed in [9] but inter-band processing is utilized at the encoder. In [6] a novel lossless compression technique is presented based on exploiting the temporal correlation under the distributed source coding paradigm. This technique operates in the pixel-domain to avoid any lossy transform and relies on syndrome decoding of trellis codes by encoding the final state of the trellis. The lattice based approach was also applied to lossless video coding, but here applying motion compensation also at the encoder side, which we want to avoid. These techniques were presented for lossless coding. In this work, instead a scalable-to-lossless distributed video codec based on Discrete Cosine Transform (DCT) is proposed based on TDWZ. This may be used for high-quality applications, where lossless is desired, but the system can not (efficiently) guarantee the resources for lossless coding. Furthermore, the coding scheme is also modified to a backward adaptive video coding system, which is evaluated to indicate the potential room for improvement in (scalable-to-lossless TDWZ) DVC.

This paper is organized as follows: Section II introduces a reversible integer DCT. The lossy TDWZ video codec is reviewed in Section III [13]. In Section IV, the TDWZ codec is modified to achieve scalable-to-lossless coding. In Section V, backward adaptive video coding is described. Finally, test conditions, results and analysis are presented in Section VI.

## II. Reversible Integer DCT

In H.264(/MPEG4 part 10) Advanced Video Coding, a $4\times4$ transform is used. It is an integer transform, but not designed for reversible transformation. A $4\times4$ transform is also a part of TDWZ coding scheme. The transforms are derived from DCT. For this transform the basic functions are cosines, and therefore the transform values and transform results are not integer. We shall apply a separable two-dimensional transform defined by one-dimensional (1-D) transforms. The initial one-dimensional transform with kernel

$$K_{DCT} = \begin{pmatrix} 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.6533 & 0.2706 & -0.2706 & -0.6533 \\ 0.5000 & -0.5000 & -0.5000 & 0.5000 \\ 0.2706 & -0.6533 & 0.6533 & -0.2706 \end{pmatrix}$$

can be implemented in a reversible version using fixed point arithmetic, but this solution leads to additional bit depth (and thereby bitplanes), which will negatively influence lossless compression. The main features of reversible integer transforms proposed and analyzed in [10] and [11] are 1)

*MMSP'10, October 4-6, 2010, Saint-Malo, France.*
*978-1-4244-8112-5/10/$26.00* 

X. Huang, A. Ukhanova, A. Veselov, S. Forchhammer, M. Gilmutdinov, "Scalable-to-Lossless Transform Domain Distributed Video Coding", *Multimedia Signal Processing (MMSP)*, St.-Malo, France, 2010.

reversibility, 2) a good approximation preserving the main transform properties and 3) limiting the number of required bitplanes.

The reversibility is achieved by using results of general matrix factorization theory. In [10] triangular elementary reversible matrix (TERM) and single-row elementary reversible matrix (SERM) are used for DCT kernel factorization. The reversible transform used in this paper is obtained by DCT kernel $K_{DCT}$ factorization via the PLUS method [11]:

$$K_{DCT} = P * L * U * S \quad (1)$$

where $S$ is Single-row, $U$ — Upper, $L$ — Low triangular elementary reversible matrices. $P$ is permutable matrix. Values of matrices can be found in [11].

It is important to maintain the order of operations in order to preserve reversibility properties. Input vector $\bar{x}$ should be multiplied by matrix-by-matrix with rounding for each intermediate coefficient:

$$\bar{y} = P * round(L * round(U * round(S * \bar{x}))) \quad (2)$$

Here operation of rounding is denoted by $round(.)$. So far as separation property is preserved for reversible transform, 2D transform can be obtained by independent applying Eq. 2 to each row and column of $4 \times 4$ input matrix $X$. The operation is denoted as:

$$\hat{Y} = [[PLUS * X] * (PLUS)^T] \quad (3)$$

To verify that this approximation leads to a transform preserving the main properties of the DCT, the variance of the differences between the DCT output and the results of the reversible integer transform were estimated. Variances were calculated for the Y component of the first 50 frames of the Foreman (CIF) sequence as follows. Each frame was divided in $4 \times 4$ non-overlapped blocks. The original (float point) DCT with the $K_{DCT}$ kernel and the reversible integer transform were applied to the $4 \times 4$ blocks $B$, thereafter the differences were calculated for each position of the transform.

$$d^B(i,j) = y^B(i,j) - \hat{y}^B(i,j),$$

where $i, j \in \{0, ..., 3\}$ are the index row and column, $y^B(i,j)$ is the element of float point matrix of DCT coefficients, obtained by:

$$Y = K_{DCT} * X * (K_{DCT})^T.$$

After that all differences corresponding to each position $(i,j)$ were calculated over all blocks $B$ and the variance value for each position was calculated:

$$VAR_D = \begin{pmatrix} 0.3887 & 0.3079 & 0.3276 & 0.3967 \\ 0.3164 & 0.2352 & 0.2776 & 0.3984 \\ 0.3403 & 0.2591 & 0.2975 & 0.4235 \\ 0.4248 & 0.3333 & 0.3595 & 0.4989 \end{pmatrix}.$$

Here the coefficient at position $(i,j)$ contains the variance value of the sequence corresponding to this position. Standard deviation for the worst case (lower right corner) is less than one. It means that energy redistribution for reversible transform case is not significant.

To examine the last feature, the ranges of one-dimensional DCT with kernel $K_{DCT}$ and its reversible approximation used in the proposed TDWZ scheme are determined. Minimal and maximal values of each DCT coefficient can be easily found from the scalar product of input vector and corresponding basis vector:

$$y(i) = \sum_{j=0}^{3} x(j) * k_{DCT}(i,j), i = 0, ..., 3,$$

where $x(j)$ are input vector values, $k_{DCT}(i,j)$ are coefficients of $K_{DCT}$. Minimal and maximal values are given in Table I. For the reversible integer transform $\bar{y}$ values are calculated by Eq. 2. In the 2D case, the number of bitplanes given by the full range for each transform coefficient is therefore 10.

TABLE I
VALUE RANGES FOR DCT4 AND REVERSIBLE TRANSFORM (1D CASE, 8-BIT INPUT)

| **Output Value** | **DCT** | | **Reversible Integer Transform** | | |
|---|---|---|---|---|---|
| | Min | Max | Min | Max | Number of bitplanes |
| y(0) | 0 | 510 | 0 | 510 | 9 |
| y(1) | -235.59 | 235.59 | -235 | 236 | 9 |
| y(2) | -255 | 255 | -255 | 255 | 9 |
| y(3) | -235.59 | 235.59 | -235 | 236 | 9 |

## III. TRANSFORM DOMAIN WYNER-ZIV VIDEO CODING

The architecture of the TDWZ video codec with feedback channel [13], which we shall base the new lossless version on is depicted in Fig. 1. It follows the same architecture as the one developed by Stanford group [1] and later further developed by the DISCOVER project [5]. However, an advanced Overlapped Block Motion Compensation (OBMC) based side information generation method [12] and an improved adaptive noise model [13] are adopted in the TDWZ video codec. These improvements lead to state-of-the-art TDWZ performance. A fixed Group of Pictures (GOP=*N*) is utilized with periodically designating one frame out of *N* in the video sequence as a key frame and utilizing these for coding the intermediate frames as Wyner-Ziv frames. The key frames are intra coded by using a conventional video coding solution with low complexity such as H.264/AVC Intra, while the Wyner-Ziv frames in between are coded using a Wyner-Ziv approach. At the encoder, Wyner-Ziv frames are partitioned into non-overlapped $4 \times 4$ blocks and applying a transform to each block. The transform coefficients within a given band $b_k, k \in \{0...15\}$, are grouped together and then quantized. DC coefficients and AC coefficients are uniformly scalar quantized and deadzone quantized, respectively. Within each coefficient band $b_k$, eight pre-defined quantization levels ($2^{M_{b_k}}$) are used, as in [5], depending on the target quality of the Wyner-Ziv frame. The quantized coefficients are then decomposed into bitplanes with bit depth $M_{b_k}$. Each bitplane is fed to a rate-compatible Low-Density-Parity-Check Accumulate (LDPCA) encoder [14] starting from the most significant bitplane. For each encoded bitplane, the corresponding accumulated syndrome is stored in a buffer together with an 8-bit Cyclic Redundancy Check (CRC). The amount of bits to be transmitted depends on the requests made by the decoder through a feedback channel as shown in Fig. 1.

At the decoder, a side information frame *Y* is interpolated and an estimated noise residue *R* is generated by using two

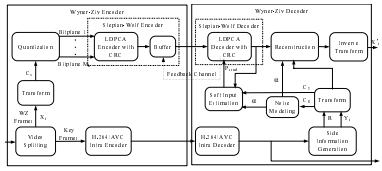



Fig. 1. Architecture of feedback channel based Transform Domain Wyner-Ziv video codec

previously decoded frames [12]. The noise residue *R* and side information *Y* undergoes the same $4\times4$ transform to obtain the transformed coefficients $C_Y$ and $C_R$, respectively. $C_R$ is used to estimate the noise distribution between the corresponding bands of the side information frame and the original Wyner-Ziv frame. Using a modeled noise distribution [13] (with estimated Laplacian parameter $\alpha$), the coefficient values of the side information frame $C_Y$ and the previous successfully decoded bitplanes, soft-input information (conditional bit probabilities $P_{cond}$) for each bitplane is estimated. With this soft-input information, $P_{cond}$, the LDPCA decoder starts to process the various bitplanes to correct the bit estimation errors. Convergence is tested by an 8-bit CRC sum and the Hamming distance between the received syndrome and the one obtained from the decoded bitplane [15]. If both the Hamming distance and CRC sum are satisfied, convergence is declared. If not, more syndrome bits are requested and decoding and testing is run again. After a bitplane is successfully decoded, a quantization interval can be obtained. It indicates the range of the original Wyner-Ziv coefficient $C_X$. Together with side information coefficients $C_Y$, noise distribution parameter $\alpha$ and the interval information, decoded coefficients within band $b_k$ of the Wyner-Ziv frame are reconstructed as in [16]. Finally, the inverse transform is performed to obtain the reconstructed Wyner-Ziv frame $X'_i$.

## IV. Scalable-to-Lossless TDWZ

In order to achieve scalable-to-lossless TDWZ video coding, some aspects of a basic lossy TDWZ video codec are required to be updated and modified. At the encoder side, the modifications are mainly on key frame coding, the transform and not applying quantization besides the quantization implicit in coding bitplanes. The key frames are lossless encoded. Obviously any lossless encoder may be applied, we chose JPEG-LS. The main modification is to apply the $4 \times 4$ reversible integer transform presented in Section II to the Wyner-Ziv frames to obtain the integer reversible transform coefficients. we shall refer to a lossy TDWZ scheme using the reversible integer DCT by rTDWZ. The coefficients are coded in a slightly modified sign-magnitude representation (where the 0 interval is associated with the positive values). Instead of pre-defined quantization matrices as in [5], the maximum absolute magnitude of transformed coefficients, $m_k$ in each band $b_k$ is calculated at the encoder. The smallest value of $M_{b_k}$, such that $2^{M_{b_k}}$ can cover the range of values, is chosen to decompose the transformed coefficient into bitplanes with the image and band dependent, minimum required bit depth $M_{b_k}$ for lossless coding. For 8 bit pixel values, $M_{b_k}$ will not exceed 10 bits, as shown in Sec. II. The deadzone quantization of the TDWZ has to be dealt with for lossless coding. No dead-zone is applied in defining the bitplanes. Actually deadzone quantization could be applied for lossy coding and then resolved by sending additional bits in the scalable-to-lossless refinement. Starting from the most significant bitplane, each bitplane is fed to a LDPCA encoder together with the CRC. The encoded bitplane is saved in a buffer and the amount of transmitted bits depends on the requests from decoder. Since all the encoded bitplanes are availabe at the encoder buffer, by controlling the number of transmitted bitplanes, the quality of transmitted and decoded Wyner-Ziv frames can be vary and be scalable-to-lossless. The Rate-Distortion (RD) performance of this scalable-to-lossless TDWZ video coding can be influenced and optimized by selective ordering of the different bitplanes and frequency bands. However, the RD optimization is not considered in the scalable-to-lossless TDWZ codec described in this paper. The decoding order is starting from the most significant bitplane to the least significant bitplane and from the low frequency band to high frequency band.

Besides employing a reversible integer transform, the modification of the decoder is mainly in the reconstruction module (Fig. 1). The rest of the decoding of Wyner-Ziv frames is the same as in the TDWZ decoder. For a given coefficient band $b_k$, if the current bitplane $M_i$ is not the final bitplane providing the required bit depth, $M_{b_k}$, i.e. $M_i < M_{b_k}$, reconstruction by the method in [16] is employed to guarantee that the reconstructed coefficients are located in the correct interval. If the current bitplane $M_i$ is the final bitplane providing the required bit depth $M_{b_k}$, all the available bitplanes generated at the encoder have been received and thus the decoded coefficient provides the lossless reconstructed transformed coefficients. After all the transform coefficients are obtained, inverse reversible integer transform is performed to reconstruct the Wyner-Ziv frame.

## V. Scalable-to-Lossless Backward Adaptive Coding

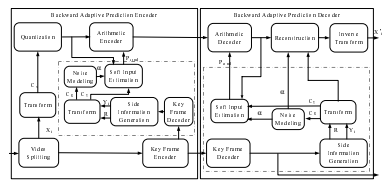



Fig. 2. Architecture of backward adaptive video coding

Interpreting the Wyner-Ziv theorem in terms of a practical Wyner-Ziv video codec as outlined above, it should be possible to achieve a RD performance similar to that of a conventional video codec under certain conditions [3]. However, based on

previous results e.g. in [12][13], there is still a gap between the performance of practical TDWZ video codec and conventional hybrid video codec (e.g. H.264/AVC Inter coding). This gap depends on the video characteristics and it may be substantial. This loss of performance of practical TDWZ video codecs may be introduced by the low quality of side information frame, an inaccurate noise model and loss of performance in the LDPCA codec etc. In order to evaluate the performance of the LDPCA codec [14] in a practical TDWZ video codec, a backward adaptive prediction coding scheme is described in this section.

Deviating from the distributed encoding, the LDPCA decoder may be replace by an arithmetic decoder using the same conditional probabilities, $P_{cond}$, as input. Now the encoder performs the same processing achieving the same conditional probabilities, $P_{cond}$, and uses these as input to the arithmetic encoder. As shown in Fig. 2, the general architecture of backward adaptive decoding scheme is the same as in the scalable-to-lossless TDWZ video codec. However, backward frame prediction (here in B-frames) is allowed to calculate the same conditional probabilities, $P_{cond}$, at encoder side. This is based on employing both the side information generation method [12] and the noise modeling module [13] also at the encoder side, hence the estimated soft input $P_{cond}$ otherwise fed into the LDPCA decoder in Wyner-Ziv coding is now available both at the encoder and the decoder and used to drive the arithmetic encoding and decoding, respectively. What we have is more like a conventional coding scheme, but now with backward adaptive motion compensation without explicit encoding of motion vectors. For ease of calculation in the experiments, we estimate the code length of the arithmetic encoder by calculating an Ideal Code Length (ICL) based on soft input $P_{cond}$, which the TDWZ utilizes in the LDPCA decoder.

For one bitplane $x$, the ICL is given by

$$L(x) = \sum_{j=0}^{n} -\log P_{cond}(x_j) \tag{4}$$

where $x_j \in \{0, 1\}$ and $P_{cond}(x_j)$ represents the estimated conditional probability of $x_j$, i.e. the symbol with index $j$. The decoder is able to lossless decode each bitplane with the same side information generation method [12], noise model [13] and the received coding bits. It is well known that context adaptive arithmetic coding can provide code lengths very close to the ICL. For both context adaptive coding and distributed source coding the ICL can take the place of an (upper bound of) the conditional entropy $H(X|Y)$, which is theoretically achievable asymptotically. The backward adaptive coding scheme is utilized to indicate the potential room for improvement of the TDWZ video codec if an ideal Slepian-Wolf codec is employed. But it may also serve as a codec in itself and evaluate the performance of modifying a DCT based video coding as H.264 to a scalable-to-lossless DCT based video coder.

## VI. Experimental Results

Performance of scalable-to-lossless TDWZ video coding is evaluated in this section. Furthermore, the performance of backward adaptive coding with ICL is also reported to illustrate potential for improvement of scalable-to-lossless TDWZ. The test sequences are *Coastguard*, *Hall Monitor* and *Foreman*, at QCIF, 15 frames per second (fps). *Hall Monitor* is dominated by a static background and thus simple capture cross-frame correlation. *Coastguard* is also characterized by in some sense simple apparent motion, due to a panning camera, while the motion of the water is not so simple. *Foreman* is a typical video telephony scene including some complex motion and scene change. Commonly used GOP size *N* equal to 2 is chosen.

Initially, the influence of introducing a reversible integer DCT is examined. The RD performance of lossy TDWZ video codec with reversible integer DCT (denoted as rTDWZ) is evaluated in Figs. 3 and 4. The performance is evaluated based on the luminance components of all the frames of a sequence. Key frames are encoded with H.264/AVC Intra [17] with the same Quantization Parameters (QP) as in [15]. For comparison, the performance of the lossy TDWZ video codec (Sec. III) and the benchmark codecs with relevant low encoding complexity, i.e. not using motion estimation at the encoder, are also reported (Figs. 3 and 4). These are H.264/AVC Intra codec, H.264/AVC No Motion codec and DISCOVER video codec. It may be noted that lossy TDWZ video coding gives better RD performance than H.264/AVC Intra codec and DISCOVER video codec on both sequences. With the reversible integer DCT, minor (acceptable) performance loss is introduced in rTDWZ, but the performance is still better than that of the DISCOVER codec.

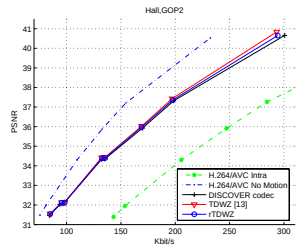



Fig. 3. Rate-Distortion performance comparison on Hall Monitor for all frames

Thereafter, the RD performance of scalable-to-lossless TDWZ video coding described in Section IV is compared with scalable-to-lossless solutions based on the JPEG 2000 codec in Figs. 5 and 6. For GOP size 2, the performance is evaluated on the luminance component of even/WZ frames only, since key frames are lossless coded. As we focus on high-quality and scalable-to-lossless, the distortion is here expressed

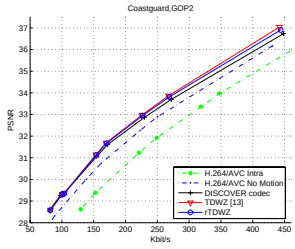



Fig. 4. Rate-Distortion performance comparison on Coastguard for all frames

by $\log(1 + MSE)$, where *MSE* is the standard mean square error. (It may be noted that values of 1 and 6 corresponds to a PSNR of 48.1 dB and 30.1 dB, respectively.) For fair comparison, both Intra frame lossless coding scheme and a low complexity Inter frame coding scheme (i.e. JPEG 2000 Diff) are included. JPEG 2000 Diff denotes compression of difference frames with JPEG 2000 coding. The difference frame $D$ is obtained by directly calculating the difference between the current frame $X_i$ and the previous key frame $X_{i-1}$.

$$D = (X_i - X_{i-1} + 128) \mod 256 \quad (5)$$

As shown in Figs. 5 and 6, the performance of the proposed scalable-to-lossless TDWZ video codec is better than JPEG 2000 Intra frame coding. Compared with low-complexity Inter frame coding schemes, the performance of scalable-to-lossless TDWZ video codec is better than JPEG 2000 Inter frame coding for the sequence *Coastguard* and it gives a comparable performance in the mostly static sequence *Hall Monitor*. Among all the RD curves, the backward adaptive coding measured by ICL always gives the best results. This indicates that the side information and noise modeling are indeed efficient, but there is still room for improvement of the scalable-to-lossless TDWZ video coding and it further shows that good performance may be achieved by video codecs using a reversible integer DCT.

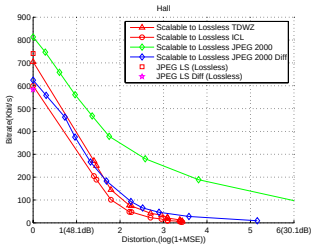



Fig. 5. Scalable-to-Lossless Rate-Distortion performance comparison on Hall Monitor for even/WZ frames only

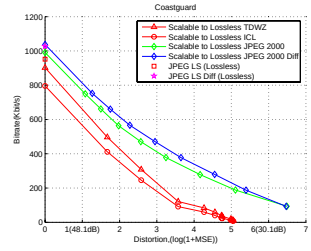



Fig. 6. Scalable-to-Lossless Rate-Distortion performance comparison on Coastguard for even/WZ frames only

Finally, the performance of different lossless coding schemes are listed and compared in Table. II and the RD performance of the scalable-to-lossless TDWZ video codec is compared with non-scalable video codecs in Figs. 7 and 8. The DCT based H.264 does not provide lossless encoding, using the JM H.264 reference software [17] but achieving high PSNR values is possible. For lossless mode using predictive coding another available H.264 codec (x.264[18]) was used. It can be seen that the lossless TDWZ video coding can save around $5\%$ bit rate compared to JPEG LS and compared to JPEG 2000 and lossless H.264 Intra frame coding the reduction is $8\%$-$13\%$ for sequences with low motion. For the sequence with intensive motion, e.g. *Foreman*, the performance of lossless TDWZ video coding is competitive as well.

Compared with low complexity Inter frame coding schemes (i.e. JPEG-LS Diff and JPEG 2000 Diff), lossless TDWZ coding gives better performance in *Coastguard* and comparable result in *Foreman* but not as good results in the almost static sequence *Hall Monitor*. Compared with lossless backward adaptive coding with ICL, there is a penalty about $13\%$-$18\%$ introduced by the practical lossless TDWZ video coding. This suggests improving the performance of the LDPCA codec employed, a topic we leave as an area for future research. Very good performance is achieved by the x.264 lossless inter, but this applies motion estimation at the encoder. Furthermore, it can be seen from the Table II that the lossless coding performance of backward adaptive coding using arithmetic coding (but here evaluated by ICL) can match conventional predictive video coding (i.e. x.264 Inter frame coding) in most cases, while providing scalable-to-lossless at the same time.

As shown in Figs. 7 and 8, the RD performance of the scalable-to-lossless TDWZ video codec is compared with non-scalable H.264/AVC (JM H.264 [17]) and JPEG LS near lossless coding. It shows that scalable-to-lossless TDWZ video can outperform H.264/AVC Intra coding and JPEG LS near lossless Intra frame coding. Compared with H.264/AVC no motion (key frames are near lossless coded with PSNR value around 80 dB) and JPEG LS near lossless Inter frame coding (key frames are lossless coded), the performance of scalable-to-lossless TDWZ video codec is better for the sequence

*Coastguard* but not as good results (especially at high quality) in the almost static sequence *Hall Monitor*. It may be noted from Fig. 7 that RD performance of JPEG LS near lossless Inter frame coding is better than scalable-to-lossless TDWZ only above 43 dB. H.264/AVC No Motion Inter frame coding is better but only gives the maximum PSNR around 65 dB.

TABLE II
LOSSLESS COMPRESSION. COMPARISON OF AVERAGE BPP OF EVEN/WZ FRAMES (GOP 2)

| | Hall Monitor | Coastguard | Foreman |
|---|---|---|---|
| JPEG-LS | 3.8987 bpp | 5.0067 bpp | 4.3499 bpp |
| JPEG-LS Diff | 3.0684 bpp | 5.4030 bpp | 4.3652 bpp |
| JPEG 2000 | 4.2798 bpp | 5.2126 bpp | 4.5991 bpp |
| JPEG 2000 Diff | 3.2852 bpp | 5.4588 bpp | 4.5454 bpp |
| x.264 Lossless Intra | 4.1446 bpp | 5.1815 bpp | 4.6621 bpp |
| x.264 Lossless Inter | 3.1162 bpp | 4.1545 bpp | 3.3324 bpp |
| Lossless TDWZ | 3.7080 bpp | 4.7463 bpp | 4.5850 bpp |
| Lossless ICL | 3.1545 bpp | 4.1880 bpp | 4.0793 bpp |

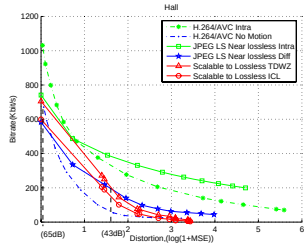



Fig. 7. Non-scalable video codecs compared with scalable-to-lossless TDWZ on Hall Monitor for even/WZ frames only

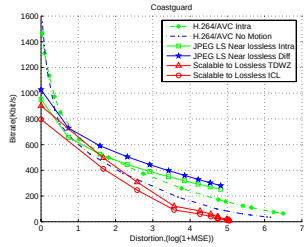



Fig. 8. Non-scalable video codecs compared with scalable-to-lossless TDWZ on Coastguard for even/WZ frames only

## VII. CONCLUSION

Scalable-to-lossless DVC was introduced based on using a reversible integer DCT as the transform in a TDWZ scheme. Experimental results show that the proposed scalable-to-lossless TDWZ video codec achieves good performance at high quality and competitive lossless performance on the test images. The codec outperformed lossless coding based on the standardized JPEG 2000. For lossless coding efficiency, the proposed scalable-to-lossless TDWZ video codec can save up to 5%-13% bits compared to lossless coding by JPEG LS, JPEG 2000 and H.264 Intra frame coding. Compared with low complexity Inter frame lossless coding schemes (i.e. JPEG-LS Diff and JPEG 2000 Diff), the proposed scalable-to-lossless TDWZ video codec gives better performance for *Coastguard*, comparable result in *Foreman* but worse results for the mostly static *Hall Monitor* sequence. Furthermore, a system based on backward adaptive coding is also introduced and tested. The results illustrate that there are still room for improvement of the scalable-to-lossless TDWZ video codec. It also showed that efficient scalable-to-lossless coding using a reversible integer DCT is feasible.

## REFERENCES


[1] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform domain wyner-ziv codec for video," in *SPIE VCIP*, Jan. 2004.

[2] J. Slepian, and J. Wolf, "Noiseless coding of correlated information sources," in *IEEE Trans. on Inform. Theory*, vol. 19, pp.471–480, July 1973.

[3] A.D. Wyner, and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," in *IEEE Trans. on Inform. Theory*, vol. 22, pp.1–10, Jan. 1976.

[4] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," in *IEEE Trans. On Image Proc.*, vol. 16, no. 10, pp.1–13, Oct. 2007.

[5] DISCOVER Project available on: www.discovervdc.org.

[6] M. Ali, and M. Murshed, "Lossless Video Coding Using Lattice Based Distributed Source Coding Techniques," in *IEEE Int'l Conf. on Video and Signal Based Surveillance*,Nov. 2006.

[7] E. Baccaglini, M. Barni, L. Capobianco, A. Garzelli, E. Magli, F. Nencini, and R. Vitulli "Low-complexity lossless compression of hyperspectral images using scalar coset codes," in *Picture Coding Symposium*, Dec. 2007.

[8] E. Magli, M. Barni, A. Abrardo, and M. Grangetto "Distributed source coding techniques for lossless compression of hyperspectral images," in *EURASIP Journal on Applied Signal Processing*, vol. 1, Dec. 2007.

[9] N.M. Cheung, C. Tang, A .Ortega, and C. S. Raghavendra "Efficient wavelet-based predictive Slepian-Wolf coding for hyperspectral imagery," in *EURASIP Journal on Signal Processing - Special Issue on Distributed Source Coding*, pp. 3180–3195, Nov. 2006.

[10] P. Hao, and Q. Shi, "Matrix factorizations for reversible integer mapping," in *IEEE Trans. Signal Processing*, vol. 49, pp.2314–2324, Oct. 2001.

[11] L. Wang, J. Wu, L.C. Jiao L. Zhang and G.M. Shi, "Lossy to lossless image compression based on reversible integer DCT," in *IEEE Int'l. Conf. on Image Processing*, pp.1037–1040, Oct. 2008.

[12] X. Huang and S. Forchhammer, "Improved side information generation for distributed video coding," in *IEEE Int'l Workshop Multimedia Signal Proc.*, Oct. 2008, pp. 223–228.

[13] X. Huang and S. Forchhammer, "Improved virtual channel noise model for transform domain wyner-ziv video coding," in *IEEE Int'l Conf. on Acoustics, Speech, and Signal Proc.*, April 2009, pp. 921–924.

[14] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, vol. 86, pp. 3123–3130, Nov. 2006.

[15] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The discover codec: architecture, techniques and evaluation," in *Picture Coding Symposium*, Nov. 2007.

[16] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in wyner-ziv video coding with multiple side information," in *IEEE Int'l Workshop Multimedia Signal Proc.*, Oct. 2007.

[17] Joint Video Team (JVT) Reference Software. Available on: http://iphome.hhi.de/suehring/tml/index.htm.

[18] H.264 Lossless video codec. Available on: http://www.videolan.org/developers/x264.html.

# ENCODER POWER CONSUMPTION COMPARISON OF DISTRIBUTED VIDEO CODEC AND H.264/AVC IN LOW-COMPLEXITY MODE

*Anna Ukhanova* [1], *Eugeniy Belyaev* [2] *and Søren Forchhammer* [1]

[1]Technical University of Denmark, DTU Fotonik, B. 343, 2800 Lyngby, Denmark
[2] Saint-Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 14 line V.O. 39, St. Petersburg, Russia

***Abstract:*** **This paper presents a power consumption comparison of a novel approach to video compression based on distributed video coding (DVC) and widely used video compression based on H.264/AVC standard. We have used a low-complexity configuration for H.264/AVC codec. It is well-known that motion estimation (ME) and CABAC entropy coder consume much power so we eliminate ME from the codec and use CAVLC instead of CABAC. Some investigations show that low-complexity DVC outperforms other algorithms in terms of encoder side energy consumption . However, estimations of power consumption for H.264/AVC and DVC stated in this paper show that for current implementations of DVC these statements could be disputed from a power consumption/compression efficiency point of view when comparing to compression algorithms based on differential frame coding (with zero search radius for ME).**

## I. INTRODUCTION

During the last years of video codec development more and more attention has been paid to the low-complexity codecs, as they are considered now for use in wireless sensor networks and another systems, where it is necessary to decrease encoding power consumption so that they can achieve longer working time, and the decoder power consumption for these systems is not an important issue. Therefore, power consumption on the encoder side has become one of the most important issues along with compression efficiency.

Distributed Video Coding (DVC) [1], [2] is a new video coding paradigm which fully or partly exploits the video redundancy at the decoder and not anymore at the encoder as in the predictive video coding, thereby shifting computation power from encoder to decoder.

Existing DVC implementations are based on the idea that the encoder uses Intra-coding part from traditional compression algorithms [3] and replaces inter coding with distributed coding. This makes the codec architecture more complex, and hence the encoder power consumption consists of two components: Wyner-Ziv encoder and Intra-encoder.

As an alternative architecture we take the baseline profile of H.264/AVC standard [3] working in differential frame coding mode (no motion coding with zero search radius for motion estimation) and using CAVLC as an entropy coder. One of the ways to show the computational performance of DVC encoder is to measure the working time on test video sequences. In particular, in [4] the working time of JM codec [5] is compared to the DISCOVER codec [6]. However, this comparison methodology should be considered to be preliminary. Firstly, the JM source code is not optimized from the time consuming point of view. Secondly, this comparison for software is not correct for hardware implementation of video compression algorithms.

Therefore, this paper proposes to use as a comparison criteria the power that the encoder consumes to provide the given peak signal-to-noise ratio (PSNR) value. Based on this criteria, we consider efficiency comparison of DVC and H.264/AVC with no motion coding for test video sequences and evaluate PSNR vs. power consumption.

Taking into account the fact that DVC codec and H.264/AVC use the same discrete cosine transform and similar quantization, and deriving simple analytical model we focus on power consumption and complexity gain comparison on the CAVLC and LDPCA encoder parts based on information about their power consumption for specific implementations, namely a $0.18\mu$ TSMC cell library, as presented in [7] and [8], respectively. This approach does not allow to estimate the overall power consumption, but gives us information about relative power consumption for a fixed PSNR value. In addition we provide an equation that explains the dependency of the power consumption gain on implementation efficiency of CAVLC and LDPC and coding efficiency of the compared algorithms.

The hypothesis of this paper is that although DVC is considered to be low-power approach for video encoding [9], common approach based on H.264/AVC with no motion coding can achieve a comparable power consumption for a given PSNR value.

The rest of the paper is organized as follows. Section II describes the H.264/AVC and DVC encoding algorithms. Section III introduces assumptions that lay a basis for de-

A. Ukhanova, E. Belyaev, S. Forchhammer, "Encoder power consumption comparison of Distributed Video Codec and H.264/AVC in low-complexity mode", *Int'l Conf. on Software, Telecom. and Computer Networks (SoftCOM)*, Bol, Croatia, 2010.

riving a simple model of power consumption of H.264/AVC and DVC and on the basis of this model a formula for power consumption gain is given. In Section IV the results of the comparison are presented.

## II. H.264/AVC AND DVC ENCODERS DESCRIPTION

### II-A. H.264/AVC encoder

This paper considers only low-complexity low-power consumption implementation of H.264/AVC. The aim of this work is to show that even this simple version of H.264/AVC can achieve good results that can compete with the results of DVC solutions. Therefore, we eliminate ME (that consumes much power [10], [11]) and apply CAVLC instead of CABAC and use H.264/AVC with no motion coding. The encoding process for each frame for H.264/AVC [3] for our differential frame coding mode includes the following operations:

1) A fixed Group of Pictures (GOP) is divided into 2 kinds of the frames, i.e. Intra-predicted (I) and Bidirectionally-predicted (B). Each frame is further divided into non-overlapping blocks of size $16 \times 16$ (macroblocks).
2) For each macroblock in I-frame Intra-prediction is performed. Then $4 \times 4$ DCT and quantization are perfomed on the residual data that is further entropy encoded with Context-Adaptive Variable Length Coder (CAVLC).
3) For each macroblock in B-frame Inter-prediction is performed. In order to avoid complex motion estimation in H.264/AVC Inter mode, differential frame coding is used. As for Intra-mode, according to the coding procedure of H.264/AVC, the produced residue is transformed, quantized and entropy coded.

### II-B. DVC encoder

This paper considers feedback channel based transform domain Wyner-Ziv video coding [12]. The encoding procedure includes the following main operations:

1) A fixed Group of Pictures (GOP=*N*) is adopted to split video sequences into two kinds of frames, i.e. Key frames and Wyner-Ziv frames. Periodically one frame out of *N* in the video sequence is named as key frame and intermediate frames are Wyner-Ziv frames. The key frames are Intra coded by using a conventional video coding solution such as H.264/AVC Intra (see Sect. II-A) while the Wyner-Ziv frames are coded using a Wyner-Ziv video coding approach.
2) Each Wyner-Ziv frame $X_i$ is partitioned into non-overlapped $4 \times 4$ blocks and a DCT [3] is applied to each of them.
3) The transform coefficients within a given band $b_k, k \in \{0...15\}$, are grouped together and then quantized. DC coefficients are uniformly scalar quantized and AC coefficients are dead zone quantized, respectively.
4) After quantization, the coefficients are binarized. The binary bits with the same significance are formed to a bitplane, which is given to a rate compatible Low Density Parity Check Accumulate (LDPCA) encoder [13]. Starting from the most significant bitplane, each bitplane is independently encoded by the LDPCA encoder, the corresponding accumulated syndrome is stored in a buffer together with an 8-bit Cyclic Redundancy Check (CRC).

## III. H.264/AVC AND DVC POWER CONSUMPTION

Taking into account, that the considered H.264/AVC scheme is not using motion estimation and preprocessing, assume that power consumption of H.264/AVC encoder depends on DCT/quantization part and entropy encoder part only. Consider that power consumption of the part that performs transform and quantization $f_{tran}^{h264}$ depends on the number of pixels, processed per time unit, and CAVLC power consumption $f_{\mathrm{CAVLC}}$ is a function of the output bitrate. Therefore analytically power consumption of H.264/AVC encoder can be written as

$$P_{h264} = f_{tran}^{h264}(F_I + F_B, W, H) + f_{\mathrm{CAVLC}}(R_I + R_B), \quad (1)$$

where $F_{\mathrm{I}}$ and $F_{\mathrm{B}}$ are frame rate for I and B frames respectively, $W$ and $H$ are frame width and height, $R_{\mathrm{I}}$ and $R_{\mathrm{B}}$ are bit rate of compressed video stream correspondent to I and B frames.

For the sake of simplicity, assume that CAVLC power consumption is a linear function of the output bitrate [14]:

$$f_{\mathrm{CAVLC}}(R_I + R_B) = (R_{\mathrm{I}} + R_{\mathrm{B}}) \cdot C_{\mathrm{CAVLC}} + P_{\mathrm{CAVLC}}^0, \quad (2)$$

where $C_{\mathrm{CAVLC}}$ is the CAVLC complexity, which depends on the concrete hardware implementation of H.264/AVC encoder, $P_{\mathrm{CAVLC}}^0$ is constant component of power consumtion.

In turn, DVC encoder power consumption consists of two parts: Intra-encoder and Wyner-Ziv encoder. Therefore, in the same way, power consumption of DVC encoder can be written as

$$\begin{aligned} P_{\mathrm{DVC}} &= f_{tran}^{h264}(F_I, W, H) + f_{\mathrm{CAVLC}}(R_I) \\ &+ f_{tran}^{dvc}(F_W, W, H) + f_{\mathrm{LDPC}}(R_W), \end{aligned} \quad (3)$$

where $F_{\mathrm{W}}$ and $R_{\mathrm{W}}$ are frame rate and bitrate of the compressed video stream for Wyner-Ziv frames, respectively.

In the same way, power consumption of LDPC part can be written as

$$f_{\mathrm{LDPC}}(R_W) = R_W \cdot C_{\mathrm{LDPC}} + P_{\mathrm{LDPC}}^0, \quad (4)$$

where $C_{\mathrm{LDPC}}$ is the LDPC complexity, which depends on the concrete hardware implementation of the LDPC encoder.

As stated above, DVC encoder and H.264/AVC use the same DCT and similar quantization part. Therefore, if $F_B = F_W$ then

$$f_{tran}^{h264}(F_I+F_B,W,H) \approx f_{tran}^{h264}(F_I,W,H)+f_{tran}^{dvc}(F_W,W,H), \quad (5)$$

From (1), (3) and (5) it follows that the power consumption gain of DVC encoder can be estimated by measurements of CAVLC and LPDC (here we assume that the power consumption estimation of LDPC is similar to LDPCA used in DVC).

Let us assume that H.264/AVC and DVC encoders use the same numbers of I frames in GOP. Then power consumption gain of DVC scheme can be written as

$$\Delta P = P_{h264} - P_{dvc} \approx f_{\text{CAVLC}}(R_B) - f_{\text{LDPC}}(R_W) = \\ = R_{\text{B}} \cdot C_{\text{CAVLC}} - R_{\text{W}} \cdot C_{\text{LDPC}} + \Delta P^0. \quad (6)$$

As was already stated above, this paper proposes to use as comparison criteria the power consumption that the encoder needs to provide a given frame distortion value. For analytical description of this criteria lwe use the following operational rate-distortion function model [15]:

$$D(R) = \frac{\theta}{R - R^0} + D^0, \quad (7)$$

where $\theta$, $R^0$ and $D^0$ are model parameters [15], $D$ and $R$ are distortion and bit rate for given video sequence.

From (6) and (7) it follows that the power consumption gain of DVC encoder for given distortion value $D(R_B^*) = D(R_W^*) = D$ can be written as:

$$\Delta P = \Big(\frac{\theta_B}{D - D_B^0} + R_B^0\Big) \cdot C_{\text{CAVLC}} - \\ -\Big(\frac{\theta_W}{D - D_W^0} + R_W^0\Big) \cdot C_{\text{LDPC}} + \Delta P^0. \quad (8)$$

Let us assume that $D_W^0 \approx D_B^0$, then (8) can be simplified as:

$$\Delta P \approx \frac{\alpha_1}{D + \alpha_2} + \alpha_3, \quad (9)$$

where $\alpha_1$–$\alpha_3$ are constants.

## IV. PERFORMANCE COMPARISON

For practical experiments the JM codec v.16.2 [5] which is the H.264/AVC reference software and DISCOVER codec [6] which is a reference software for distributed video coding were used. Practical results were obtained for four test video sequences "foreman" and "hall monitor" at QCIF (176x144) resolution, 15 Hz, 150 frames and "foreman" and "hall monitor" at CIF (352x288) resolution, 30 Hz, 300 frames.

For power consumption estimation of CAVLC and LDPC we have used the results published in [7] and [8], respectively, and the linear models (2) and (4) were used to extract the power consumption values. All power consumption measurements were achieved for $0.18\mu$ TSMC cell library. Although this library could not be considered very new, both CAVLC and LDPC results were obtained for the same experimantal conditions and, therefore, if one solution consumes less power on $0.18\mu$ TSMC cell library than another, most likely it will also consume less power on the advanced technologies.

The rate-distortion performance (Fig. 1–4) and the power consumption estimated as described in Section III. The results were combined to evaluate PSNR/Power consumption. Figures 5–8 show the relative power consumption for different GOP sizes. Thereafter, we chose the points on the curves with minimum power consumption for DVC and H.264/AVC and calculate the difference. This difference is shown on Fig. 9–12 denoted as "estimated results". These graphs also show the results for the proposed model (9), that approximates the DVC power consumption gain relative to H.264/AVC in low-complexity mode very accurately.

Results show that for most relevant visual quality range (30-40 dB) DVC with LDPC allows to decrease power consumption of entropy coding about 15-60% compared to CAVLC for H.264/AVC in differential frame coding mode. Taking DCT and Quantization blocks into account this gain is not dramatic and may not warrant the use of feedback in the DVC scheme considered.

## V. CONCLUSION

In this paper two low-complexity codecs were discussed and compared. We have proposed a simple analytical model that allows to estimate power consumption gain of DVC from H.264/AVC. This model shows the dependency between encoding algorithms efficiency and implementation efficiency of CAVLC and LDPC and power consumption gain.

In this paper we have compared only the kernel of the video codecs, and the difference is rather small. If we consider other costs, the difference may be insignificant. Therefore, taking into account the complexity increase due to DVC introducing a different coding unit, than the one already used for intra/keyframes and the small gain in comparison to H.264/AVC may be preferable to use H.264/AVC in differential coding mode for the systems that require low encoding complexity.

The authors would like to note that this paper uses simple models and assumptions and focused on LDPC as replacement of CAVLC. This does not take into consideration some other components of the power consumption like memory access, residual data calculation for Inter-mode in H.264/AVC and bit-plane coding along with CRC calculation in DVC. The model is planned to be made more precise in the future work.

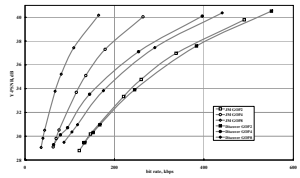


**Fig. 1**. Rate-distortion performance comparison for “hall monitor qcif“

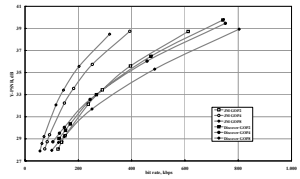


**Fig. 2**. Rate-distortion performance comparison for “foreman monitor qcif“

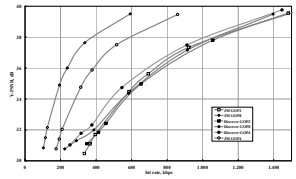


**Fig. 3**. Rate-distortion performance comparison for “hall cif“

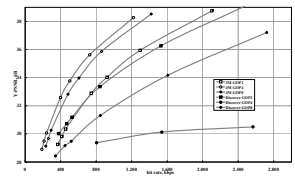


**Fig. 4**. Rate-distortion performance comparison for “foreman cif“

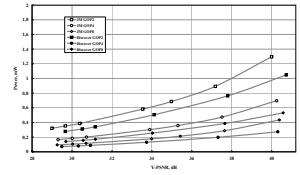


**Fig. 5**. Relative power consumption comparison for “hall qcif“

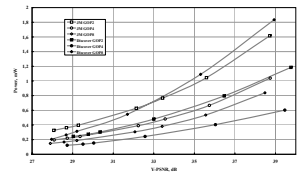


**Fig. 6**. Relative power consumption comparison for “foreman qcif“

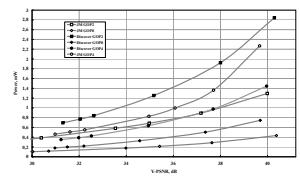


**Fig. 7**. Relative power consumption comparison for “hall cif“

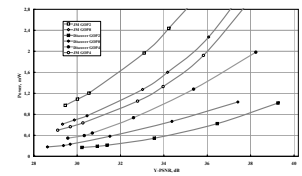


**Fig. 8**. Relative power consumption comparison for “foreman cif“

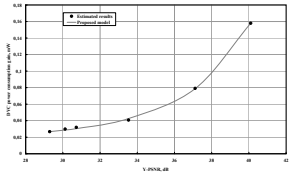



**Fig. 9**. DVC power consumption gain for "hall qcif"

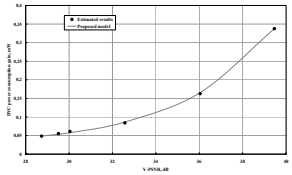



**Fig. 10**. DVC power consumption gain for "foreman qcif"

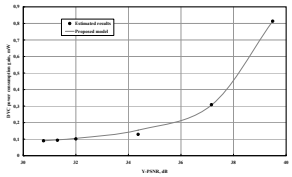



**Fig. 11**. DVC power consumption gain for "hall cif"

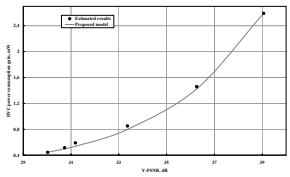



**Fig. 12**. DVC power consumption gain for "foreman cif"

## VI. REFERENCES


[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. "Distributed video coding", *Proc. of IEEE, Special issue on advances in video coding and delivery*, vol. 93, no. 1, pp. 71-83, Jan. 2005.

[2] A. Aaron, R. Zhang, and B. Girod. "Wyner-Ziv coding of motion video", *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 240-244, Nov. 2002.

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003.

[4] P. L. Dragotti and M. Gastpar, Distributed Source Coding: Theory, Algorithms and Applications. *Academic Press*, ISBN-13: 978-0123744852, 2009.

[5] H.264/AVC JM Reference Software, available on: http://iphome.hhi.de/

[6] DISCOVER codec, available on: http://discoverdvc.org/

[7] F. Huang, S. Lei, "A High Performance and Low Cost Entropy Encoder for H.264 AVC Baseline Entropy Coding", *ICCCAS*, 2008.

[8] R. Swamy, S. Bates and T. L. Brandon, "Architectures for ASIC implementations of low-density parity-check convolutional encoders and decoders",*ISCAS* (5), pp. 4513-4516, 2005.

[9] A. Aaron and B. Girod, "Wyner-Ziv video coding with low-encoder complexity," Proc. Picture Coding Symposium, *PCS*, San Francisco, CA, December 2004. Invited paper.

[10] V. Iverson, J. McVeigh, B. Reese, "Real-time H.24-AVC codec on Intel architectures ", *International Conference on Image Processing (ICIP)*, 2004

[11] M. Chien, J. Huang, P. Chang, "Complexity Control for H.264 Video Encoding over Power-Scalable Embedded Systems", *The 13th IEEE International Symposium on Consumer Electronics (ISCE)*, 2009.

[12] A. Aaron, R. Zhang, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *In Proc. SPIE Visual Com. and Img. Proc.*, vol. 5308, pp. 520-528, January 2004.

[13] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *EURASIP Signal Process. Journal, Special Section on Distributed Source Coding*, vol. 86, pp. 3123-3130, Nov. 2006.

[14] X. Lu, Y. Wang, and E. Erkip, "Power efficient H.263 video transmission over wireless channels," *International Conference on Image Processing*, vol. 1, pp. 533-536, 2002.

[15] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels", *IEEE Journal on Selected Areas in Communications*, vol.18, pp. 10121032, 2000.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

# Low-latency video transmission over high-speed WPANs based on low-power compression

Eugeniy Belyaev[1], Andrey Turlikov[2] and Anna Ukhanova[3]
[1]Intel Corporation, Intel Labs, Saint-Petersburg, Russia
[2]State University of Aerospace Instrumentation, Saint-Petersburg, Russia
[3]DTU Fotonik, Technical University of Denmark

***Abstract*—This paper disscuses latency-constrained video transmission over high-speed wireless personal area networks. Low-power single-layer video compression is proposed as an alternative to others video processing approaches. End-to-end distortion and end-to end latency in video transmission system are analyzed. A near-optimal video source rate control based on MINMAX quality criteria is introduced. Practical results for video encoder based on H.264/AVC standard are also given.**

## I. Introduction

In the last few years a number of high throughput wireless personal area networks (WPANs) have appeared, such as the IEEE 802.15.3c standard [1]. These networks have low power data transmitters and throughput up to 6 Gigabit per second. Therefore, they can provide a transmission of high definition video from mobile device to display over wireless channel instead wired cable. These networks are lacking of the best method for the choice of video processing so this task still remains actual.

To choose the best video processing approach for such type of networks the following restrictions and requirements should be taken into account. To provide low power consumption at mobile transmitter one-pass low-complexity and low-memory approaches without motion compensation or temporal filtering (intra processing only) have to be used. At the same time these solutions should provide very low transmission latency, continuous video playback at the receiver and acceptable visual quality for all variety of video sources: sequences of computer graphics, snapshots, natural and mixed images.

From our point of view there are four potential approaches that can satisfy these restrictions and requirements:

- uncompressed video transmission;
- intra single-layer video compression;
- intra scalable video coding;
- distributive video coding.

Let us consider the list stated above. The straightforward solution in these networks can be based on uncompressed video transmission [1]. This approach does not require any compression algorithm and provides low processing latency. On the other hand, several disadvantages arise. Firstly, the throughput of a wireless channel is time-varying where, in addition, other traffic such as audio or IP data can be transmitted along with video data. Therefore, it could not be guaranteed that the channel rate is high enough for continuous video playback. Secondly, this approach does not use the channel in an efficient way in the sense of throughput and energy consumption, because it does not take into account the video source redundancy. In the third place, this solution requires technical change at the network layers like combination of video data unequal error protection and special automatic repeat request methods [1]. Fourthly, the pixel partitioning technique which is used for unequal error protection is not efficient in rate-distortion sense especially for desktop snapshot type of the video that contains a lot of details commensurable with pixel size.

The second solution can be based on Scalable Video Coding (SVC) and unequal error protection of different quality layers [2]. In this case latency-constrained video transmission over variable wireless channels is achieved due to dropping the higher enhancement layers of the scalable video [3]. However, known SVC algorithms have higher computation complexity and lower compression efficiency than single-layer video compression [4].

The third solution can be based on Distributive Video Coding (DVC) [5] which became very popular in the last few years. Practical results show that DVC can provide compression efficiency better than single-layer video compression in intra-mode [6]. Many papers present DVC as very low-power approach for video compression. But at this time there are no DVC implementations which verify it. In addition DVC encoder has to contain two compression cores: Winer-Ziv encoder and traditional intra-encoder, and has very high power consumption on the decoder side that is not efficient for consumer products.

And final solution can be based on single-layer video compression which is most extended in video processing devices. This approach has low encoder and decoder complexity, it has good compression efficiency in rate-distortion sense and do not require any changes at link and physical network layers.

The particular properties of the video processing methods are stated in Table I. Complex comparison shows that intra single-layer video compression is more preferable for video data transmission over high-speed wireless personal area networks.

Note, that transmission system based on solutions, which were described above, has a set of parameters (like quantization step, macroblock type, transport packet length, modulation and code scheme and so on) which have to change in real time depending on video source and wireless channel

978-1-4244-6398-5/10/$26.00 ©2010 IEEE

E. Belyaev, A.Turlikov, A. Ukhanova, "Low-latency video transmission over high-speed WPANs based on low-power compression", *IEEE Wireless Communications & Networking Conference (WCNC)*, Sydney, Australia, 2010.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

Table I
VIDEO PROCESSING APPROACHES COMPARISON

| Video processing approach | Encoder / Decoder complexity | Compr. efficiency | Is network layers modification needed? |
|---|---|---|---|
| Uncompressed video transmission | very low / very low | very low | yes |
| Intra single-layer video compression | low / low | medium | no |
| Intra scalable video coding | medium / medium | low | partly |
| Distributive video coding | low / high | high | no |

states. In common case, algorithms, which are controlling these parameters, should minimize video distortion based on objective or subjective quality criteria taking into account power consumption and transmission latency restrictions. In common case it is very difficult task and an open problem at the current time.

This work is a continuation of our research in this way. In papers [11], [12] we propose latency-constrained video source rate control algorithm based on MINMAX quality criteria for transmission of video data over constant throughput channels. In this paper we extend our approach taking into account power consumption restrictions and variable throughput channel and adopt it to intra single-layer video compression based on H.264/AVC standard [7]. We introduce several assumptions, formulate optimization task and propose corresponding one-pass video source rate control algorithm.

This paper is organized as follows. Section II describes low-power implementation of H.264/AVC Encoder. In Sections III–IV end-to-end distortion and end-to-end latency in video compression and transmission systems are discussed. In Section V video source rate control based on MINMAX quality criteria is introduced. Finally, the practical results for different test video sequences are shown.

## II. Low-power video compression based on H.264/AVC Standard

H.264/AVC compression standard [7] is based on exploiting the spatial and temporal redundancy of video sources. This is achieved by using motion estimation and compensation, intra-frame prediction, discrete cosine transform, quantization, entropy coding and others methods.

To achieve low-power compression, low computation complexity and memory consumption is needed. To decrease the memory size it is proposed to eliminate motion estimation and to use intra-coding only. In this case the encoder can be implemented by using internal memory which is needed to store 32 pixel lines of the input video only. For example, for the resolution size of $1920\times1080$ 180 Kbytes is needed instead of more than 6 Mbytes in the motion compensation case.

In addition to the proposed scheme, even for this small memory size the simple case of the temporal redundancy removal could be used. It is often the case that many regions in the current and the previous frames are the same in computer graphics and desktop snapshots (static regions). Therefore, it is possible to detect "static" macroblocks at the encoder side by calculating hash function value. If hash function value for the current macroblock is equal to the corresponding hash function value for the previous frame, it can be encoded in SKIP mode and the decoder shows the corresponding macroblock which was transmitted earlier.

The further decrease of the computational complexity can be achieved by using DC intra-prediction, $4\times4$ DCT and CAVLC (Context-adaptive variable-length coding) compression modes only. For improving the encoding performance and achieving absolutely RGB-lossless compression it is proposed to use reversible YCoCg 4:4:4 color space transform.

## III. End-to-end distortion in transmission system

Let us assume that each video frame is separated into non-overlapping *units* that include several macroblocks. The end-to-end distortion $d_t$ for unit $t$ in wireless video communication systems consists of two main components [13]:

$$d_t = d(q_t) + d_c, \tag{1}$$

where $d(q_t)$ is distortion caused by quantization at the encoder side and $d_c$ is distortion caused by channel errors and error concealment algorithm at the decoder side. In this paper we describe video transmission based on MINMAX quality criteria [14] that can be interpreted as follows. For each unit $t$ the distortion $d_t$ should be provided, so that

$$\textit{minimize} \max_t d_t. \tag{2}$$

Usually the automatic repeat request (ARQ) method is used to achieve reliable data transmission over an unreliable channel. For each packet the receiver sends to transmitter special message (acknowledgement) that indicates if the packet is received correctly or not. If the packet is not received correctly then the transmitter sends it again. The probability of this situation can be defined as $p_t = 1-(1-p_b)^l$, where $p_b$ is bit error rate (BER) and $l$ is transport packet length assuming independent bit errors. If the packet is not received after $n$ retransmissions then it is dropped at the transmitter side with the probability $p_t^n$ and the decoder shows the corresponding co-located macroblocks which were transmitted earlier.

Assume that *channel rate controller* chooses modulation and coding scheme (MCS) and transport packet size to maximize the throughput depending on channel feedback. In this case the transmission scheme that provides the BER $< 10^{-4}$ is chosen. Then the optimal transport packet size for this BER values can be chosen [16] to guarantee that packet loss probability is $p_t^n < 10^{-10}$. Therefore, for further optimization we can disregard packet losses. It means that for end-to-end distortion minimization it is enough to minimize quantization distortion $d(q_t)$ only.

## IV. End-to-end latency in transmission system

### A. Video transmission system description

The video transmission system model discussed in this paper is shown in the Figure 1. Consider the system timing

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

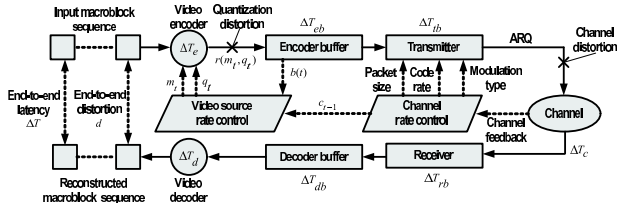



Figure 1. End-to-end distortion and end-to-end latency in video transmission system

is discrete and slotted. The slot time is a part of the system time $[t, t+1)$ and time moment $t$ refers to the end of this slot. *Channel rate controller* chooses the transmission scheme that maximizes the channel throughput. Taking into account transport packet headers, ARQ and time division between different types of traffic, let us define channel throughput for video data $c_t$ as the number of bits that are transmitted during time slot $t$. The video source gives encoder a *unit* that contains $M$ macroblocks of the encoded frame. After compressing unit $t$ into $r(q_t)$ bits, where $q_t$ is a quantization step, encoder places it into the *encoder buffer*. Depending on the number of bits in the encoder buffer and channel state, *video source rate controller* chooses the quantization step $q_t$ and macroblocks type $m_t \in \{intra, skip\}$ for the next unit.

The number of bits in the encoder buffer $b^e(t)$ after placing there a new compressed unit $t$ and transmitting over the channel with the throughput $c_t$, changes as follows:

$$b^e(t) = \max\{0, b^e(t-1) - c_t\} + r_t(q_t, m_t). \quad (3)$$

Data on the receiver side is accumulated for some time $L$ after which the decoding and playing starts.

### B. Latency definition, necessary and sufficient conditions

Generally, latency $\Delta T$ between the time moment when some unit has been sent to the encoder and the time moment when this unit has been shown at the receiver device display consists of the following components:

$$\Delta T = \Delta T_e + \Delta T_{eb} + \Delta T_{tb} + \Delta T_c + \Delta T_{rb} + \Delta T_{db} + \Delta T_d, \quad (4)$$

where $\Delta T_e$ and $\Delta T_d$ are the encoding and decoding processing latency, $\Delta T_{eb}$ is the encoder buffer latency, $\Delta T_{tb}$ is transmitter buffer latency, $\Delta T_{rb}$ is receiver buffer latency, $\Delta T_{db}$ is the decoder buffer latency, $\Delta T_c$ is the channel transmission latency.

Let us suppose that the encoder and the decoder work real-time and values $\Delta T_e$, $\Delta T_d$ and $\Delta T_{tb}$, $\Delta T_{rb}$ are significantly less than $L$. In [8] it was shown that

$$\Delta T_{eb} + \Delta T_c + \Delta T_{db} = L, \quad (5)$$

if the number of bits in the encoder buffer is

$$b^e(t) \le b_{eff}(t) = \sum_{i=t+1}^{t+L\cdot f\cdot N} c_i, \quad (6)$$

where $b_{eff}(t)$ is the *effective buffer size* [15], $N$ is a number of units in the frame and $f$ is a frame rate.

For time-varying wireless channel $b_{eff}(t)$ is equal to the sum of the future channel rates in time interval $[t, ..., t+L]$ and it can not be calculated at the time moment $t$, because future channel rates are not known yet. Therefore, effective buffer size is usually estimated at the encoder side by using channel model [9]. However, for time-varying wireless channel it is not possible to guarantee that estimated value $\hat{b}_{eff}(t) \le b_{eff}(t)$ for any time moment $t$. Then the situation when $b^e(t) > b_{eff}(t)$ is possible and latency requirements (5) do not hold.

### C. Required latency restoration approach

If required $L$ value is low (e.g. 1ms) then we can use the following approach to restore the required latency. At the time moment $t+L$ we can calculate effective buffer size $b_{eff}(t)$. If at the time moment $t+L$ inequality (6) does not hold then at the time moment $t$ latency requirements (5) do not hold. It means that at the time moment $t+L$ number of bits in decoder buffer is $b^d(t+L) = 0$ and decoder can not start the reconstruction process.

Assume that decoder works in real-time, therefore decoding time for INTRA unit is less than $\Delta T_d^{intra} \le \frac{1}{f \cdot N}$ and decoding time for SKIP unit is close to zero $\Delta T_d^{skip} \approx 0$. Then, to restore equation (5) we propose the following algorithm.

**Step 1.** Compress all units in SKIP mode until at the time moment $t^*$ encoder buffer will be emptied $b^e(t^*) = 0$.

**Step 2.** Compress all units in SKIP mode at the time interval $[t^*, ..., t^* + 2 \cdot L + \frac{1}{f \cdot N}]$.

At the time moment $t+L$ encoder buffer contains not more than $n = 2 \cdot L \cdot f \cdot N + 1$ units. At the time moment $t^*$ all these units will be available at the decoder side together with SKIP units that were formed in the time interval $[t+L, t^*]$. To decompress it decoder spends time

$$\Delta T_d(n) \le (2 \cdot L \cdot f \cdot N + 1) \cdot \Delta T_d^{intra} \le 2 \cdot L + \Delta T_d^{intra}. \quad (7)$$

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

It means that at the time moment $t^* + 2 \cdot L + \frac{1}{f \cdot N}$ encoder buffer will contain SKIP units and decoder buffer will be empty. This event is equivalent to the system starting state.

## V. Video source rate control algorithm

### A. MINMAX optimization task description

Note that for high-speed video transmission we can use high-resolution quantization hypothesis [10] that defines distortion as $d(q) = q^2/12$, therefore MINMAX criteria (2) corresponds to

$$\textit{minimize} \max_t q_t. \tag{8}$$

Suppose that despite statistical properties of the units in the frame may be quite different from each other, statistical properties of all frames vary insignificantly. It means that there is only one scene in the input video sequence. This assumption does not hold true generally, because video sequence usually consists of subsequences (scenes) with different statistical properties. To make understanding of the algorithm with several scenes easier, let us initially take into account the case when video sequence has only one scene.

Let us formulate rate control optimization task according to the latency requirements (6) and the MINMAX quality criteria (8). For each unit $t$ it is necessary to choose the quantization step $q_t$, so that

$$\begin{cases} minimize \max\limits_t q_t \\ b(t) \le b_{eff}(t). \end{cases} \tag{9}$$

### B. Solution of MINMAX task by consecutive search algorithm

Solution of the task (9) can be found by the following hypothetic algorithm which consists of the following two steps:

**Step 0.** (*Initialization*)
0.1 Set $\{q_i\} = \{0, 1, ..., q_{max}\}$, $i \leftarrow 0$.
0.2 Go to Step 1.1

**Step 1.**
1.1 $\tilde{q} \leftarrow q_i$, $\tilde{b}(0) \leftarrow 0$.
1.2 For units $t = 0, 1, ...$ calculate
$\tilde{b}(t) \leftarrow \max\{0, \tilde{b}(t-1) - c_t\} + r_t(\tilde{q})$.
If $\tilde{b}(t) > b_{eff}(t)$ then $i \leftarrow i + 1$ and go to Step 1.1

The algorithm described above is called the *consecutive search algorithm.*

**Theorem 1.** *Consider $\tilde{q}$ the solution found by the consecutive search algorithm. There is no sequence of quantization steps $y_1, y_2, \dots$ for which $\max\limits_t y_t < \tilde{q}$ that does not lead to effective buffer size exceeding.*

**Proof.** Suppose that *consecutive search algorithm* has stopped at the step $i$. Then for each step $j < i$ for every unit $t$ quantization step $x_t = q_j$ was chosen. From consecutive search algorithm description follows that after encoding unit $\tau$ number of bits in encoder buffer $\tilde{b}(\tau) > b_{eff}(\tau)$.

Let us choose any sequence of quantization steps $y_1, y_2, \dots$, where $y_t \le q_j$, and $b(t)$ is the number of bits in encoder buffer, when unit $t$ is encoded with $y_t$ value. Then $y_t < x_t$, consequently,

$$r(y_t) \ge r(x_t) \tag{10}$$

So if $\tilde{b}(0) = b(0) = b_0$, then from (3) and (10) follows that $\tilde{b}(t) \le b(t)$. It means that exists such $\tau' \le \tau$ that

$$b(\tau') > b_{eff}(\tau'). \blacksquare \tag{11}$$

### C. Single-scene MINMAX rate control algorithm

Consecutive search algorithm is a hypothetic one that shows the solution of (9), but can not be implemented in a real-time system, because it is impossible to rerun data transmission after effective buffer size exceeding.

Therefore, this paper proposes an algorithm that allows to find the estimation of $\tilde{q}$ for the consecutive search algorithm. Consider $\hat{q}_t$ to be the estimation of $\tilde{q}$ value. It is supposed to estimate $\tilde{q}$ value as follows. All units are compressed with quantization step $\hat{q}_t$ until the number of bits in the buffer $b^e(t)$ will not exceed effective buffer size $b_{eff}(t)$. This exceeding means that it is impossible to hold the $\hat{q}_t$ value for the given channel throughput for fixed end-to-end latency without increasing it. So, the end-to-end latency exceeds its initial value $L$ and, consequently, the required latency restoration approach is used and the estimation of $\hat{q}_t$ increases. The algorithm consists of the following three steps.

**Step 0.** (*Initialization*)
0.1 Set $\hat{q}_0 \leftarrow q_0$, $t \leftarrow 0$, $b^e(0) \leftarrow 0$.
0.2 Go to Step 1.1

**Step 1.** (*Buffer accumulation*)
1.1 $t \leftarrow t + 1$, $\hat{q}_t \leftarrow \hat{q}_{t-1}$.
1.2 $b^e(t) \leftarrow \max\{0, b^e(t-1) - c_t\}$.
1.3 Compress unit $t$ with quantization step $\hat{q}_t$.
1.4 If $b^e(t) > b_{eff}(t)$ go to Step 2.1
1.5 $b^e(t) \leftarrow b^e(t) + r(\hat{q}_t)$ and go to Step 1.1

**Step 2.** (*Latency restoration*)
2.1 Compress all units in SKIP mode until at the time moment $t^*$ encoder buffer size $b^e(t^*) = 0$.
2.2 Compress all units in SKIP mode at the time interval $[t^*, ..., t^* + 2 \cdot L \cdot f \cdot N + 1]$.
2.3 $t \leftarrow t^* + 2 \cdot L \cdot f \cdot N + 1$, $\hat{q}_t \leftarrow \hat{q}_t + \Delta q^+$ and go to Step 1.1

**Theorem 2.** *Consider that consecutive search algorithm finds the quantization step value $\tilde{q}$. Then for the proposed algorithm with initial value $\hat{q}_0 \le \tilde{q}$, the inequality $\hat{q}_t < \tilde{q} + \Delta q^+$ holds true for any time moment $t$.*

**Proof.** Let $\tilde{b}(t)$ be the buffer size for the consecutive search algorithm. From its description

$$\tilde{b}(t) \le b_{eff}(t). \tag{12}$$

Let us suppose that $\hat{q}_0 \le \tilde{q}$ and at the time moment $\tau$ this inequality holds true firstly:

$$\tilde{q} \le \hat{q}_\tau < \tilde{q} + \Delta q^+. \tag{13}$$

So at this moment the number of bits in the encoder buffer (see Step 2.1) is:

$$b^e(\tau) = 0. \tag{14}$$

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

From (13) for $t \geq \tau$ the following inequality holds true:

$$r(\hat{q}_t) \leq r_t(\tilde{q}), \tag{15}$$

so that from (3), (12), (14) and (15) follows that at the time moment $t \geq \tau$ the number of bits in the buffer is:

$$b^e(t) \leq \tilde{b}(t) \leq b_{eff}(t). \tag{16}$$

Thereby, from the time moment $\tau$ the statement of Step 1.4 of this algorithm fails. Consequently, the algorithm will not reach Step 2.3 and parameter $\hat{q}_t$ will not be increased.■

From Theorem 2 follows that after algorithm adaptation for the source and channel properties the quantization step for each unit will not exceed the quantization step for the solution of the task (9) and the value not more than $\Delta q^+$. The time of the adaptation depends on the source and channel properties, and starting value $q_0$ and $\Delta q^+$. With $\Delta q^+$ increase the time of the adaptation decrease, and vice versa.

### *D. Scene change and virtual buffer concept*

Now let us take a look at the video sequences that consist of several scenes $s_0, s_1, \ldots, s_n$. Then MINMAX optimization task (9) should be applied for each scene. Let $\tilde{q}(s_i)$ be a solution provided by consecutive search for scene $s_i$. If $\tilde{q}(s_{i+1}) \geq \tilde{q}(s_i)$, then algorithm proposed above will adapt to a new scene. However, if $\tilde{q}(s_{i+1}) < \tilde{q}(s_i)$, then algorithm will not decrease $\hat{q}_t$, that means that the quality will not be improved.

Therefore, to overcome this problem we introduce an heuristic approach based on a *virtual buffer* concept. For each unit $t$ the following value is calculated:

$$b_v^e(t) \leftarrow \begin{cases} b^e(t), \ if \ t = t^*, \\ \max\{0, b_v^e(t-1) - c_v(t)\} + r(\hat{q}_t - \Delta q^-), \\ if \ t \neq t^*, \end{cases} \tag{17}$$

where $t^*$ is a number of the first unit in the current frame, $c_v(t)$ is a virtual channel rate that is calculated as follows:

$$c_v(t) = \frac{b_{eff}^{min}(w,t)}{L \cdot f \cdot N}, \tag{18}$$

where $b_{eff}^{min}(w,t)$ is a minimum of the effective buffer size for the previous $w$ frames

$$b_{eff}^{min}(w,t) = \min_i b_{eff}(i), i \in \{t^* - w \cdot N, \ldots, t^* - 1\}. \tag{19}$$

In addition, the difference between the number of bits that is placed into the buffer and maximum number of bits that could be transmitted is accumulated:

$$\Delta r_v \leftarrow \sum_{i=t^*}^{t^*+N-1} r(\hat{q}_t - \Delta q^-) - c_v(t). \tag{20}$$

Let us take a look on the virtual buffer concept. If $\Delta r_v > 0$, the number of bits sent to the transmission buffer is more then the number of bits sent to the channel and this can lead to the effect of latency exceeding the limit during the transmission of the next frames. On the other side, the bit size distribution for units in each frame may be so, that this can happen even if $\Delta r_v \leq 0$. Therefore, in addition $b_v^e(t)$ is calculated. Thus, if before the encoding of the unit $t^*$ the following statements are fulfilled:

$$\begin{cases} \max\limits_i b_v^e(i) \leq b_{eff}^{min}(w, t^* - 1), \ i \in \{t^* - N, \ldots, t^* - 1\}, \\ \Delta r_v \leq 0, \end{cases} \tag{21}$$

and rate control was not in the latency restoration mode during coding of previous frame, the quantization step value is modified as follows:

$$\hat{q}_t \leftarrow \max\{0, \hat{q}_t - \Delta q^-\}. \tag{22}$$

### *E. Using static units detection in rate control*

For improving video quality for the low channel throughput case, static units can be transmitted repeatedly in lossless mode. However, we have to take into account that the types of the units (static or non-static) in the future are unknown. Therefore we should keep a significant part of the encoder buffer free for non-static units. Thus, if rate control works in buffer accumulation mode then lossless mode is used for static units, if they were not transmitted as lossless earlier and number of bits in encoder buffer $b^e(t) \leq \alpha \cdot b_{eff}(t)$, $\alpha \in [0, ..., 1]$.

## VI. Practical results

To obtain practical results the suggested rate control algorithm was embedded into the low-power H.264/AVC encoder that was shortly described in Section II. In the rate control algorithm the following parameters were used: $L = 1$ ms, $M = 2$, $w = 5$, $\Delta q^+ = 3$, $\Delta q^- = 2$, $\alpha = 0.15$.

Channel throughput $c_t$ simulation is executed as following. At first, propagation measurements in the presence of human activity for a 60 GHz channel [17] were used for obtaining of the temporal variations of the channel SNR($t$). Secondly, SNR($t$) vs. $p_t$ dependencies for each MCS were calculated based on transport packet length $l = 4092$, number of retransmissions $n = 10$ and SNR/BER curves from IEEE 802.15.3c standard proposals documents [18]. Finally, for each SNR($t$) value one of the MCS was chosen that provide BER$< 10^{-4}$.

The performance of the discussed algorithm was tested on two video sequences with $1920 \times 1080$ frame resolution, frame rate $f = 60$. The first test video sequence ("Breeze") is a typical movie which contains natural images. The second test video sequence ("Desktop") corresponds to computer desktop snapshots: running office applications and dragging windows. Figure 2 shows video source rate and peak signal-to-noise ratio (PSNR) for the given channel throughput. For the convenience of graphic expression PSNR$= 70$ dB corresponds to the absolutely lossless compression.

Practical results show that in good channel condition case the low-power encoder provides lossless video source rate equal to 1.5 Gbps for natural video sequences that allows to economize channel throughput or use it for other data traffic.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2010 proceedings.

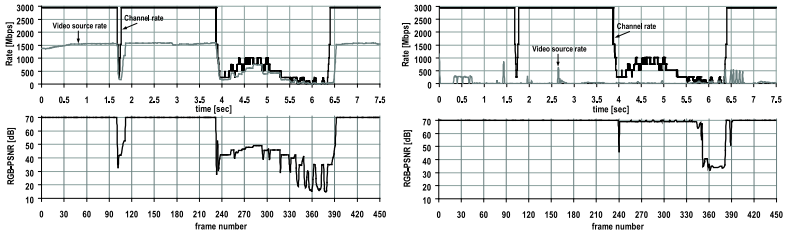



Figure 2. Practical results for "Breeze" (on the left side) and "Desktop" (on the right side) video sequences

In bad channel state the proposed rate control algorithm provides adaptation to varying channel conditions and guarantees acceptable video quality for the given channel throughput. Greater effect is obtained for desktop video sequences that contain a lot of static regions.

## VII. Conclusions

In this paper the latency-constrained video transmission over high-speed wireless personal area networks was discussed. In the complex comparison it leads to the conclusion that single-layer video compression suits at most for the described situations. The video source rate control algorithm based on MINMAX quality criteria was proposed and practical results for the real channel are shown. The proposed rate control was constructed in one-pass mode: it does not need recompression and it does not necessary need the channel model constuction. By this, the number of operations that is needed for macroblocks types and quantization steps selection in the unit is tiny in comparison to the encoder operations. Therefore, the proposed algorithm does not contribute much to the general encoder power consumtion.

The future works will be devoted to the rate-distortion performance comparison of single-layer video compression with other approaches like uncompressed video, scalable video coding and distributed video coding for high-speed wireless personal area networks taking into account transmission latency and power consumtion restrictions.

## VIII. Acknowledgements

The authors would like to thank Søren Forchhammer for his contribution to this work.

## References

[1] H. Singh, Jisung Oh, Changyeul Kweon, Xiangping Qin, Huai-Rong Shao and Chiu Ngo, "A 60 GHz wireless network for enabling uncompressed video communication", *IEEE Communications Magazine*, vol. 46, Issue 12, pp. 71 – 78, 2008.

[2] M.Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video", *IEEE Transactions on Circuits and Systems for Video Technology*, 2001.

[3] Yaser Pourmohammadi Fallah, Hassan Mansour, Salman Khan, Panos Nasiopoulos, Hussein M. Alnuweiri, "A Link Adaptation Scheme for Efficient Transmission of H.264 Scalable Video Over Multirate WLANs", *IEEE Transactions on Circuits and Systems for Video Technology*, vol.18, No.7, 2008.

[4] H. Schwarz, D.Marpe, and T.Wiegand, Overview of the Scalable Video Coding Extension of the H.264 / AVC Standard // IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, No. 9, pp. 1103-1120, 2007.

[5] T. Kuganeswaran, X. Fernando, L. Guan, "Distributed video coding and transmission over wireless fading channel", *Canadian Conference on Electrical and Computer Engineering*, 2008.

[6] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed Video Coding",*Proceedings of the IEEE*, vol. 93, No. 1,pp. 71-83, 2005.

[7] Advanced video coding for generic audiovisual services. *ITU-T Recommendation H.264 and ISO/IEC 14496-10 (AVC)*, 2009.

[8] A.R. Reibman and B.G Haskell, "Constraints on variable bit-rate video for ATM networks", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, Issue 4, pp. 361 – 372, 1992.

[9] Chi-Yuan Hsu, Antonio Ortega and Masoud Khansari, "Rate control for robust video transmission over burst-error wireless channels", *IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios*, vol. 17, pp. 756–773, 1999.

[10] H. Radha, M. Dai, D. Loguinov, "Rate-distortion modeling of scalable video coders", *International Conference on Image Processing*, pp. 1093–1096, 2004.

[11] E. Belyaev, A. Turlikov, and A. Ukhanova, "Rate-distortion control in wavelet-based video compression systems with memory restriction", *XI International Symposium on Problems of Redundancy in Information and Control Systems*, pp. 13–17., 2007.

[12] E. Belyaev, A. Dogadaev and A. Ukhanova. "MINMAX rate control in near-lossless video encoders for real-time data transmission", *XII International Symposium on Problems of Redundancy in Information and Control Systems*, pp. 3–9, 2009.

[13] Fan Zhai, Y. Eisenberg, T.N. Pappas, R. Berry, A.K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission", *IEEE Transactions on Image Processing*, vol. 15, No. 1, pp. 40 – 53, 2004.

[14] N. Cherniavsky, G. Shavit, M.F. Ringenburg, R.E. Ladner, E.A. Riskin, "Multistage: A MINMAX bit allocation algorithm for video coders", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, Is. 1, pp. 59 – 67, 2007.

[15] A. Ortega and M. Khansari, "Rate control for video coding over variable bit rate channels with applications to wireless transmission", *International Conference on Image Processing*,vol. 3, pp.338–391, 1995.

[16] M. Liinaharja, "Studies on the Performance of Some ARQ Schemes", PhD thesis, Helsinki University of Technology, 2006.

[17] S. Collonge, G. Zaharia, G. El Zein, "Influence of the human activity on wide-band characteristics of the 60GHz indoor radio channel", *IEEE Transactions on Wireless Communications*, vol.3, pp. 2396–2406, 2004.

[18] IEEE 802.15 WPAN Millimeter Wave Alternative PHY Task Group 3c (TG3c), Contributions and documents, 2009.

# Temporal Scalability Comparison of the H.264/SVC and Distributed Video Codec

Xin Huang [1], Anna Ukhanova [2], Eugeniy Belyaev [2], Søren Forchhammer [1]
[1] DTU Fotonik, Technical University of Denmark
[2] State University of Aerospace Instrumentation, Saint-Petersburg, Russia

***Abstract*— The problem of the multimedia scalable video streaming is a current topic of interest. There exist many methods for scalable video coding. This paper is focused on the scalable extension of H.264/AVC (H.264/SVC) and distributed video coding (DVC). The paper presents an efficiency comparison of SVC and DVC having reduced encoder complexity. Moreover, temporal scalability is described for these two algorithms, and it is analyzed and compared.**

## I. Introduction

Scalable video coding is very interesting for multimedia networks. Various clients might require decoding of the same video at different resolutions and qualities. Therefore, scalable coding encodes the video only once and enables decoding at different qualities, spatial and temporal resolutions. It makes scalable video coding attractive for different applications. The Moving Picture Experts Group (MPEG) has recently introduced the Scalable Video Coding (SVC) standard [1], which is an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard [2]. SVC achieves very good compression performance. On the other hand, SVC entails a higher complexity at the encoder side. Another approach is taken in the field of Distributed Video Coding [3] as a new video coding paradigm to deal with lossy source coding using side information to exploit the statistics at the decoder to reduce computational demands at the encoder. Using DVC, for example, the burden of motion estimation and compensation can be shifted from the encoder to the decoder. This implies low power / low complexity encoders.

The paradigm of distributed source coding (DSC), which has its roots in the theory of coding correlated sources developed by Slepian and Wolf [4] for the lossless case and Wyner and Ziv [5] for the lossy case, has recently become the focus of different kinds of video coding schemes [6], [7]. DVC is promising in creating reversed complexity codecs for power constrained (handheld) devices. Unlike regular broadcast oriented video codecs with high encoding complexity and low decoding complexity, reversed complexity codecs have low encoding complexity but high decoding complexity.

SVC could be used in the situation when we have many receivers and it is needed to receive the data at different bitrates. This can be used for the following:

- video transmitted over Internet for the users with different receiving rate;
- digital TV (DVB-T, DVB-H, ATSC, DTMB, ISDB, SBTVD);
- wireless transmission (on the base of Wi-MAX, WiFi).

Another case is when we have to control the transmission rate depending on the situation in the channel. If the channel becomes worse, it is possible to use scalable stream for power saving [11]. As for DVC, it will suit the situation better, when there are limitations for the complexity and memory of the encoder, and also for power consumption. In a number of resource critical applications, a complex video encoder is a disadvantage in terms of physical size and power consumption. DVC is proposed to apply in areas, where the cost of separated video encoders is the primary concern:

- wireless video surveillance;
- low-power video sensors;
- wireless digital cameras and camera embedded mobile phones.

The goal of this paper is to explore the efficiency of the temporal scalability of DVC and SVC for reduced encoder complexity. By comparing the coding performance, the advantages and disadvantages of scalable DVC and H.264/SVC are analyzed and discussed. The rest of this paper is organized as follows: Section II briefly describes different types of scalabilities in video coding. In Section III, temporal scalability in H.264/SVC is introduced. In Section IV, coding procedures of state-of-the-art DVC is described. Temporal scalablities and complexity of H.264/SVC and DVC are compared

9781–4244–3941–6/09/$25.00 

X. Huang, A. Ukhanova, E. Belyaev, S. Forchhammer, "Temporal scalability comparison of the H.264/SVC and Distributed Video Codec", *International Conference on Ultra Modern Telecommunications (ICUMT)*, St. Petersburg, Russia, 2009.

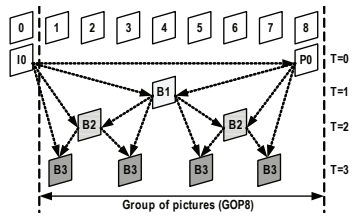



Fig. 1. Temporal scalability scheme.

in Section V.

## II. Types of scalability

Scalable extension of the H.264/AVC standard is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems. "Scalability" in this paper means removal of parts of the bit stream to adapt it to the different needs or preferences of end users as well as to the network conditions.

The main idea of scalable coding is that coder forms the bit-stream from several layers: base layer and enhancement layers. The base layer of a bit stream is always coded in compliance with a non-scalable profile of H.264/AVC (single-layer coding). For next enhancement layers encoding the previous layers (that may include the base layer) is needed. Each layer is characterized by its own bit rate and visual quality. Thus, receivers could decode the necessary layers to provide with the necessary bit rate and visual quality.

There exist different ways of the video data processing to form the streams with the properties described above:

- *Temporal scalability.*
- *Spatial scalability.*
- *SNR-scalability.*
- *Combined scalability.*

Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) or frame rate (temporal resolution), respectively. With SNR (quality) scalability, the substream provides the same spatial-temporal resolution as the complete bit stream, but with a lower fidelity where fidelity is often informally referred to as signal-to-noise ratio (SNR). The different types of scalability can also be combined, so that a multitude of representations with different spatial-temporal resolutions and bit rates can be supported within a single scalable bit stream [10]. As temporal scalability is the most obvious scalability type, we only focus on this case.

## III. Temporal scalability in H.264/SVC

Temporal scalability in H.264/SVC is achieved by using hierarchical coding structures with B-pictures [8]. The pictures of the temporal base layer are only predicted from previous pictures of this layer. The enhancement layer pictures can be bidirectionally predicted by using the two surrounding pictures of a lower temporal layer as references. A picture of the temporal base layer and all temporal refinement pictures between the base layer picture and the previous base layer picture build a group of pictures (GOP). In each GOP, the frame at the lowest level is called the key frame and it is encoded as I- or P-frames. Each temporal layer is marked by an additional identifier $T$. $T$ is equal to 0 for pictures of the temporal base layer and is increased by 1 from one temporal layer to the next.

Figure. 1 shows an example of building hierarchical B-picture structure for the case of GOP containing 8 frames. In this case base temporal layer $T = 0$ consists of only the single key (frame 8) of this GOP. Next layer $T = 1$ consists of single B-picture (frame 4) that requires two reference frames in forward and backward direction (frame 0, frame 8) from layer $T = 0$. In the same manner B-picture (frame 2) in the layer $T = 2$ also requires two reference frames (frame 0, frame 4) from layers $T = 0$ and $T = 1$ accordingly. The following steps are done in a similar manner. For the implementation of this type of scalability it is necesary to store all 8 frames in the encoder memory. This brings additional delay and increases the size of the memory used. Therefore, if it is needed to decrease the memory size and delay, temporal scalability could be used in low-delayed mode. However, this will lead to a efficiency degradation.

Temporal scalable bit-stream can be generated by using hierarchical prediction structures without any changes to H.264/MPEG4-AVC. The encoding process for each frame includes the following operations:

1) Inter-frame prediction
   - Motion estimation (4x4, 4x8, 8x4, 8x8, 8x16,16x8,16x16 inter-block search). For each block in the current frame it is necessary to make the search for the most similar block in the previous frame(s).
   - Motion compensation. This means the difference calculation between the current block and blocks found in the reference frame(s).

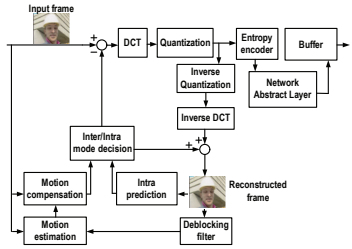



Fig. 2. Main operarions for SVC encoding [2]

2) Intra-frame prediction (DC-prediction, Vertical, Horizontal, Diagonal and others predictions). The prediction of the current block is done by the pixels of the left and upper blocks.
3) Deblocking filter. This filter is used to remove the blocking effect for the improvement of the motion compensation.
4) Discrete Cosine Transform (4x4 DCT, 8x8 DCT, 16x16 DCT).
5) Scalar quantization.
6) Intra/Inter prediction mode decision - in this stage the best prediction mode and best DCT type for current macroblock is chosen.
7) Entropy encoding
   - CABAC (Context-Adaptive Binary Arithmetic Coder);
   - CAVLC (Context-Adaptive Variable Length Coder).
8) Network Abstract Layer - it forms a H.264/SVC compatible stream which can be transmitted over any network.

The decoding process using H.264/SVC includes:

1) Entropy decoding (depending on what was used for encoding).
2) Scalar dequantization.
3) Inverse Discrete Cosine Transform.
4) Motion compensation.
5) Error resielence algorithm - describes what to do if some part of the bit stream was damaged.
6) Deblocking filter.

## IV. DVC AND ITS TEMPORAL SCALABILITY

Feedback channel based transform domain Wyner-Ziv video coding is one DVC approach. The architecture of transform domain Wyner-Ziv video codec [6] is depicted in Fig. 3. The encoding procedure includes the following main operations:

1) A fixed Group of Pictures (GOP=*N*) is adopted to split video sequences into two kinds of frames, i.e. Key frames and Wyner-Ziv frames. Periodically one frame out of $N$ in the video sequence is named as key frame and intermediate frames are WZ frames. The key frames are Intra coded by using a conventional video coding solution such as H.264/AVC Intra while the Wyner-Ziv frames are coded using a Wyner-Ziv video coding approach.
2) Each Wyner-Ziv frame $X_i$ is partitioned into non-overlapped $4 \times 4$ blocks and a DCT [2] is applied to each of them.
3) The transform coefficients within a given band $b_k, k \in \{0...15\}$, are grouped together and then quantized. DC coefficients are uniformly scalar quantized and AC coefficients are dead zone quantized, respectively.
4) After quantization, the coefficients are binarized. The binary bits with the same significance are formed to a bitplane, which is given to a rate compatible Low Density Parity Check Accumulate (LDPCA) encoder [12]. Starting from the most significant bitplane, each bitplane is independently encoded by the LDPCA encoder, the corresponding accumulated syndrome is stored in a buffer together with an 8-bit Cyclic Redundancy Check (CRC). The amount of transmitted bits depends on the requests made by the decoder through a feedback channel. Although latency is introduced by a feedback channel, encoder complexity can be minimized with this feedback channel based rate control mechanism.

The decoding procedure is described as follows:

1) A side information frame $Y_i$ and its corresponding noise residual frame $R$ are created in the side information generation module [13] by using previously decoded frames. The side information frame $Y_i$ is seen as a 'noise' version of the encoded Wyner-Ziv frame $X_i$, the estimated noise residual frame $R$ is utilized to express the correlation noise between the Wyner-Ziv frame $X_i$ and the side information frame $Y_i$.
2) The estimated noise residual frame $R$ and side information frame $Y$ undergo the DCT to obtain the coefficients $C_R$ and $C_Y$. Taking $C_R$ and $C_Y$ as inputs of a noise model module [15], the noise distribution between corresponding frequency bands

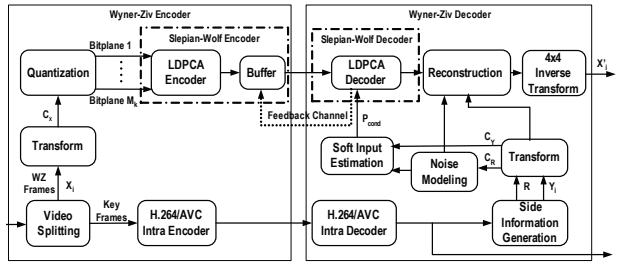



Fig. 3. Feedback channel based transform domain Wyner-Ziv video codec architecture

of the side information frame $Y_i$ and the Wyner-Ziv frame $X_i$ is modeled.

3) Using a modeled noise distribution, the coefficient values of the side information frame $C_Y$ and the previous successfully decoded bitplanes, soft-input $P_{cond}$ (conditional bit probabilities) for each bitplane is calculated.
4) With the obtained soft-input $P_{cond}$, the LDPCA decoder starts to process various bitplanes to correct bit errors. Convergence is tested by the 8-bit CRC sum and the Hamming distance. If the Hamming distance is different from zero or the CRC sum is incorrect after a certain amount of iterations, the LDPC decoder requests more accumulated syndrome bits from the encoder buffer via the feedback channel to correct the existing bit errors. If both the Hamming distance and CRC sum are satisfied, convergence is declared, guaranteeing a very low error probability for the decoded bitplane.
5) After successful LDPCA decoding, the obtained bitplanes are grouped together to form a set of decoded quantization symbols for each band $b_k$. With the received quantization information, the decoded quantized symbols are used to calculate the correct intervals in which the Wyner-Ziv coefficients are located. Together with side information coefficients $C_Y$, noise distribution parameter $\alpha$ and the interval information, decoded coefficients within band $b_k$ of the Wyner-Ziv frame are reconstructed.
6) After all the coefficients bands are reconstructed, 4×4 block inverse transform is performed to obtain the reconstructed Wyner-Ziv frame $X_i'$.

Compared with the DISCOVER DVC codec in [14], the novelty of the implemented DVC codec is combining an improved Overlapped Block Motion Compensation (OBMC) based side information generation module [13] and an adaptive virtual channel noise model module [15]. Beside the novel DVC aspects, our DVC implementation is also extended with temporal scalability in this paper according to GOP size 8 example as shown in Fig. 1. Each temporal layer in DVC can be encoded independently without storing any reference frames. The temporal layer $T = 0$ consists of the H.264/AVC Intra coded frames (frame 0 and frame 8), while the other layers are Wyner-Ziv coded frames. During the decoding, Wyner-Ziv frames in the next layer $T = 1$ (frame 4) needs two previous decoded key frames in forward and backward direction (frame 0, frame 8) from layer $T = 0$ for decoding. Similarly, Wyner-Ziv frames (frame 2) in layer $T = 2$ utilize two frames (frame 0, frame 4) from layers $T = 0$ and $T = 1$ for decoding.

Due to the feedback channel based rate control mechanism in our DVC implementation, the *coded data* (e.g. the coded frame 4) in higher layers still needs to be stored in a buffer before lower layer frames (e.g. frame 0 and frame 8) is successfully decoded. The size of *coded data* to be stored may be equivalent of up to 1.5 frames with simple rate control [16]. Ideally, if an efficient rate control is employed, it may be possible to avoid store these data for the realization of temporal scalability in DVC.

## V. Comparison of scalability performance

In order to make fair scalability comparisons between DVC and H.264/SVC, the Joint Scalable Video Model (JSVM) reference software v.9.15 [9] which has processed video stream in temporal scalable mode is used. It is important to note, that the comparison was made for reduced encoder complexity. SVC worked in the

Intra mode without memory for the frames. For this the most complicated blocks were turned off (e.g. motion estimation). In the differential coding mode the encoding complexity was also minimal but the additional memory for the frames was needed. The test conditions adopted in this paper are the DISCOVER project test conditions, commonly used in the DVC literature [13][14]. The test sequences "hall monitor" and "coastguard" are coded at QCIF, 15 frames per second (fps). The key frames are encoded using H.264/AVC Intra and the QPs are chosen so that the average PSNR (Peak Signal-to-Noise Ratio) of the WZ frames is similar to the average PSNR of the key frames. The RD performance is evaluated for the luminance component of the key frames, WZ frames and hierarchical B frames. GOP consists of 8 frames: IWWWWWWWI for Wyner-Ziv encoding and IBBBBBBBI for SVC encoding (taking into account that I frames were encoded in a similar manner). The temporal scalability results are shown in Figs. 4–9.

TABLE I
COMPLEXITY AND MEMORY SIZE COMPARISON FOR ENCODER FOR GOP8

| Encoder type | Computation complexity | Memory |
|---|---|---|
| H.264/SVC Intra | Intra prediction, DCT, Quantization, Entropy encoding, IDCT, Dequantization | Less than 1 frame |
| DVC | DCT, Quantization, LDPCA encoder, CRC | Equivalent to 1 frame |
| H.264/SVC Differential frame coding | Inter/Intra mode decision, DCT, Quantization, Entropy encoding, IDCT, Dequantization | more than 8 frames |

If there are no restrictions on the complexity of the decoder, then as shown in Figs. 4–9 the use of state-of-the-art DVC is preferable in the case when we need to have minimal memory and encoder complexity. The efficiency of state-of-the-art DVC is better than SVC for the same memory size and complexity. If the size of the memory at the encoder is not limited, the H.264/SVC has better results (see Table I). If there is a limitation on the encoder complexity then simplified H.264/SVC (e.g. without motion compensation) is better.

In this work, we evaluate temporal scalability. It is straightforward possible for our DVC scheme to provide SNR-scalability by selecting of bitplanes. Furthermore the basic layer (or key frames) could be lower resolution, thus also providing spatial scalability. The choice of scalability using DVC may be made without changing the DVC encoder, thus only the decoder needs to be modified.

## VI. CONCLUSION

The efficiency of the temporal scalability of state-of-the-art DVC and SVC with reduced complexity encoding are discussed in this paper. If there are the strong restrictions on the encoder memory and complexity then only H.264 in the Intra-frme mode can provide temporal scalability. If the encoder memory is close to one frame and we have complexity restrictions at the encoder then DVC shows better results. If there are no encoder memory restrictions, but only restriction for the complexity, it is better to use H.264 in the Differential Frame Coding mode. Thus, it is shown that with the encoder memory restrictions and availability of the temporal scalability the best method of the encoding should be chosen taking into account the memory restrictions Due to the existing performance gap, it is necessary to further improve the coding efficiency of DVC. The minimization of the encoder complexity overhead for scalable coding without sacrificing coding efficiency has become an active research area in the video coding community. As a continuation of this work in the future, additional research for spatial and SNR scalability will be conducted.

## REFERENCES


[1] International Organization for Standardization, "Introduction to SVC Extension of Advanced Video Coding", ISO/IEC JTC1/SC29/WG11, International Organization for Standardization, Coding of Moving Pictures and Audio, Pozna , Poland, July 2005. URL: http://www.chiariglione.org/mpeg/technologies/mp04-svc/svc/.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003.

[3] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed Video Coding". Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, January 2005.

[4] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," IEEE Trans on Inf. Theory, vol. 19, no. 4, pp. 471-480, Jul 1973.

[5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans on Inf. Theory, vol. 2, no. 1, pp. 1-10, Jan 1976.

[6] A. Aaron, R. Zhang, and B. Girod, "Transform-domainWyner-Ziv codec for video," In Proc. SPIE Visual Com. and Img. Proc., vol. 5308, pp. 520-528, January 2004.

[7] H. Wang, N. M. Cheung, and A. Ortega, "A framework for adaptive scalable video coding using Wyner-Ziv techniques," EURASIP Journal on Applied Signal Proc., pp. 1-18, 2006.

[8] S. Lim, J. Yang, and B. Jeon,"Fast Coding Mode Decision for Scalable Video Coding", *10th Int'l Conf. on Advanced Communication Technology*, vol. 3, pp. 1897-1900, 2008.

[9] JSVM 9.15 software package, CVS server for the JSVM software. http://iphome.hhi.de/

[10] H. Schwarz, D. Marpe, and T. Wiegand "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9, September 2007

[11] E.Belyaev, V. Grinko, A. Ukhanova, "Power Saving Control for the Mobile Receivers in the DVB-H based on the Scalable Extension of H.264/AVC Standard" *Wireless Telecommunications Symposium*, 2009

[12] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *EURASIP Signal Process. Journal, Special Section on Distributed Source Coding*, vol. 86, pp. 3123-3130, Nov. 2006.

[13] X. Huang, and S. Forchhammer, "Improved side information generation for distributed video coding," *IEEE Int'l Workshop om Multimedia Signal Processing*, pp. 223-228, Oct. 2008.

[14] Available on: www.discoverdvc.org.

[15] X. Huang, and S. Forchhammer, "Improved virtual channel noise model for transform domain Wyner-Ziv video coding," *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 921-924, April 2009.

[16] M. Morbee, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel," *IEEE Int'l Conference on Acoustics, Speech and Signal Processing* , pp. 521-524, April 2007.

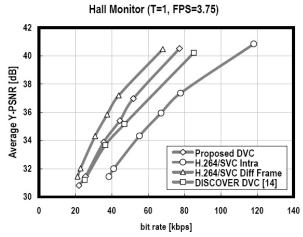



Fig. 4. RD comparison for SVC and DVC. "hall" (T = 1)

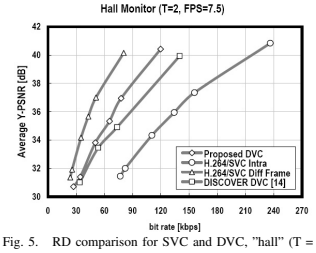



Fig. 5. RD comparison for SVC and DVC, "hall" (T = 2)

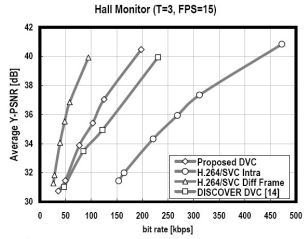



Fig. 6. RD comparison for SVC and DVC. "hall" (T = 3)

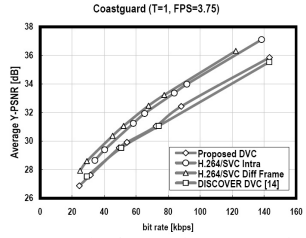



Fig. 7. RD comparison for SVC and DVC. "coastguard" (T = 1)

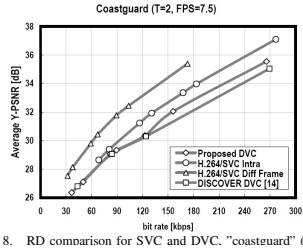



Fig. 8. RD comparison for SVC and DVC. "coastguard" (T = 2)

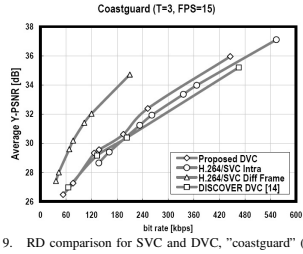



Fig. 9. RD comparison for SVC and DVC, "coastguard" (T = 3)

# Appendix B

# Performance results for Rate-Distortion-Complexity optimization

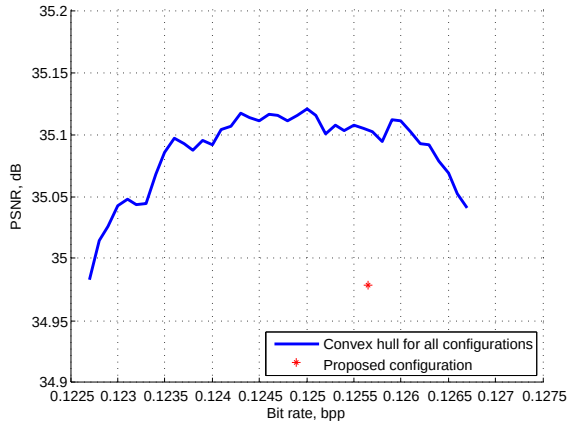



**Figure B.1:** Performance results for the proposed solution compared to offline optimization for "Crew" with 60% complexity

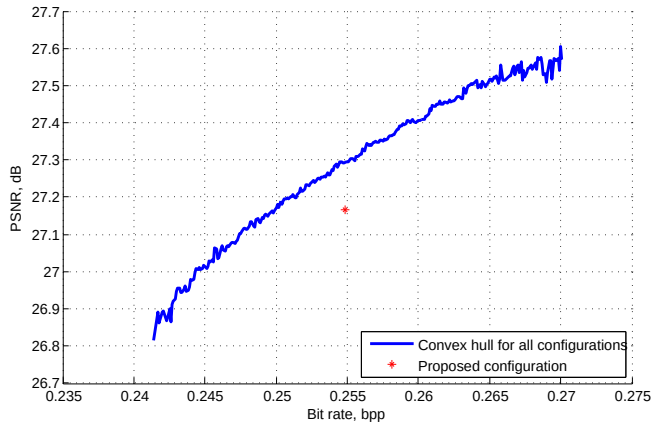



**Figure B.2:** Performance results for the proposed solution compared to offline optimization for "Mobile" with 60% complexity

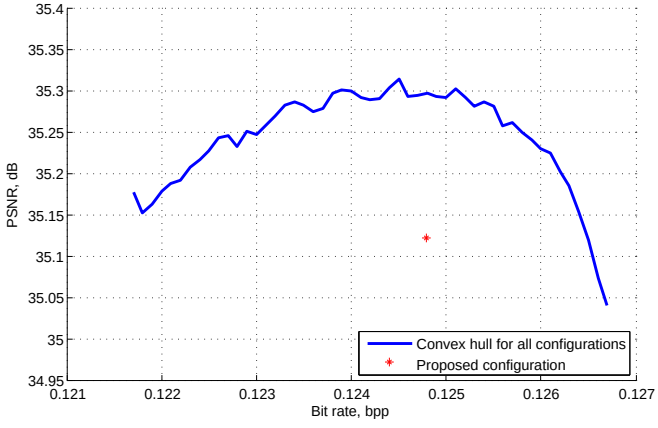



**Figure B.3:** Performance results for the proposed solution compared to offline optimization for "Crew" with 70% complexity

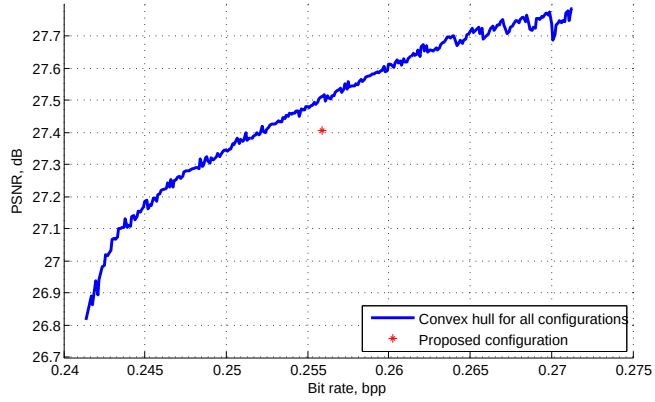



**Figure B.4:** Performance results for the proposed solution compared to offline optimization for "Mobile" with 70% complexity

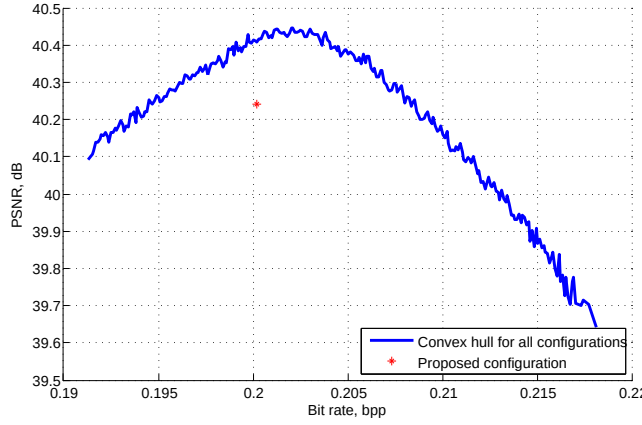



**Figure B.5:** Performance results for the proposed solution compared to offline optimization for "News" with 70% complexity

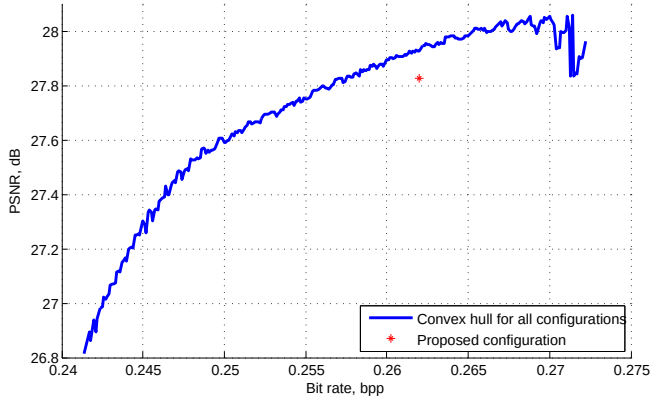



**Figure B.6:** Performance results for the proposed solution compared to offline optimization for "Mobile" with 80% complexity

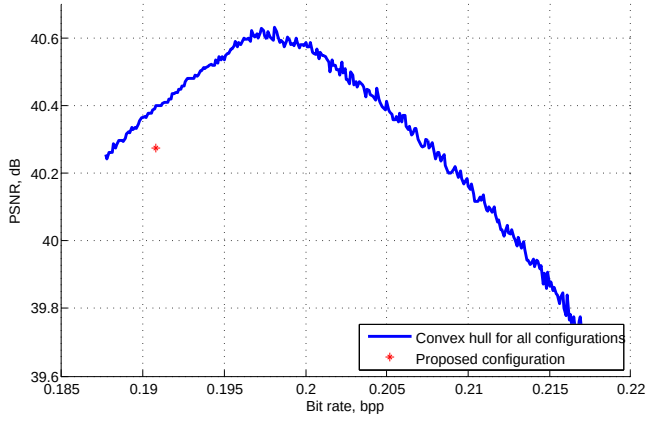



**Figure B.7:** Performance results for the proposed solution compared to offline optimization for "News" with 80% complexity

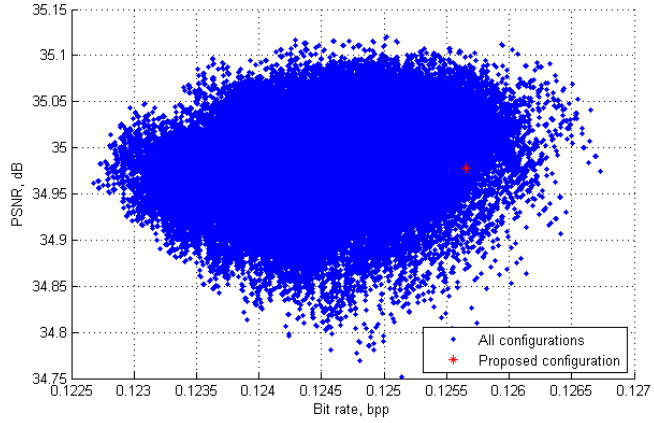



**Figure B.8:** Performance of different configurations for the sequence "Crew", 60% complexity

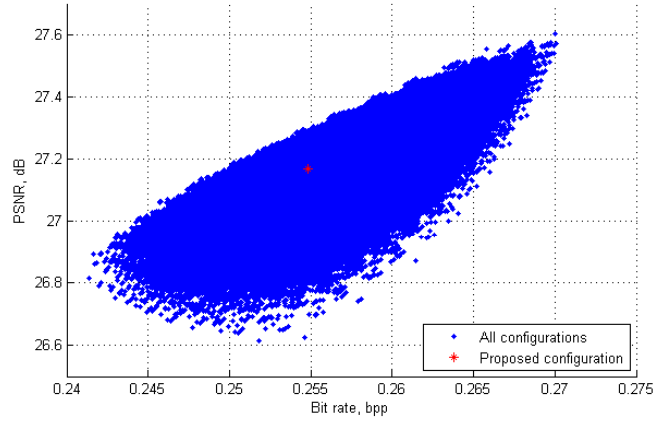



**Figure B.9:** Performance of different configurations for the sequence "Mobile", 60% complexity

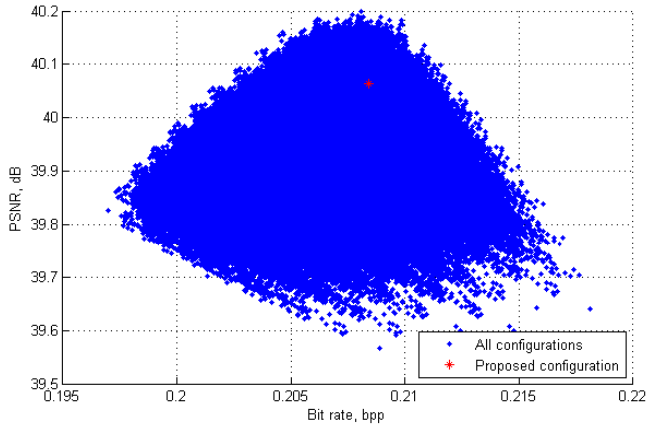



**Figure B.10:** Performance of different configurations for the sequence "News", 60% complexity

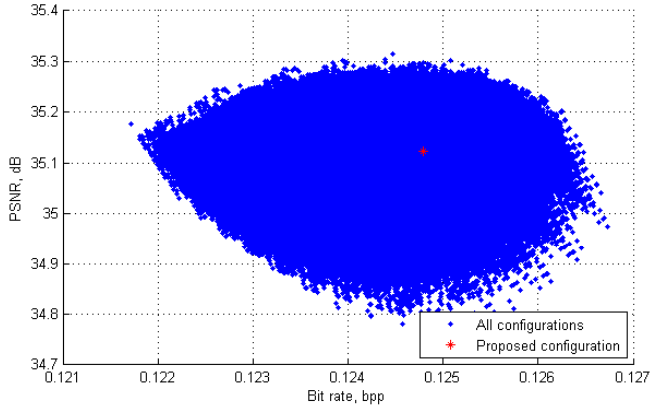



**Figure B.11:** Performance of different configurations for the sequence "Crew", 70% complexity

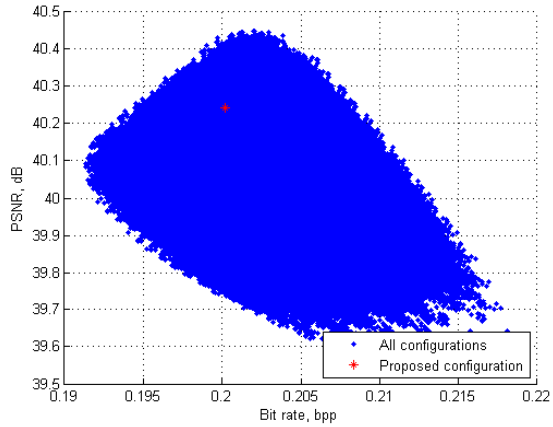



**Figure B.12:** Performance of different configurations for the sequence "News", 70% complexity

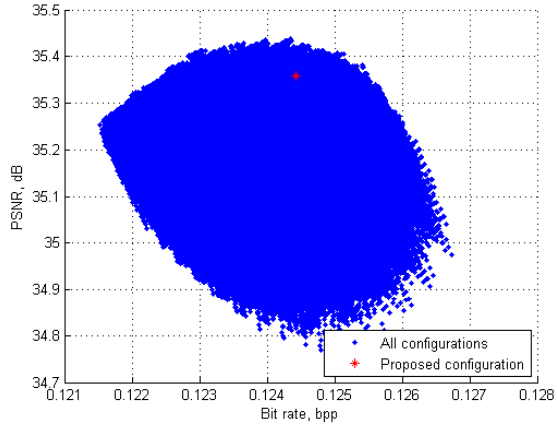



**Figure B.13:** Performance of different configurations for the sequence "Crew", 80% complexity

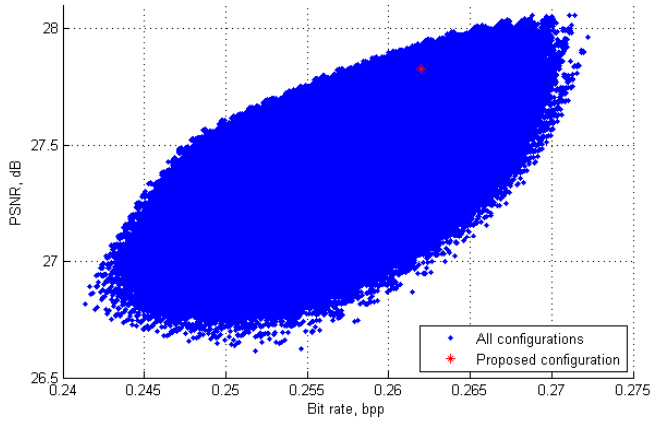



**Figure B.14:** Performance of different configurations for the sequence "Mobile", 80% complexity

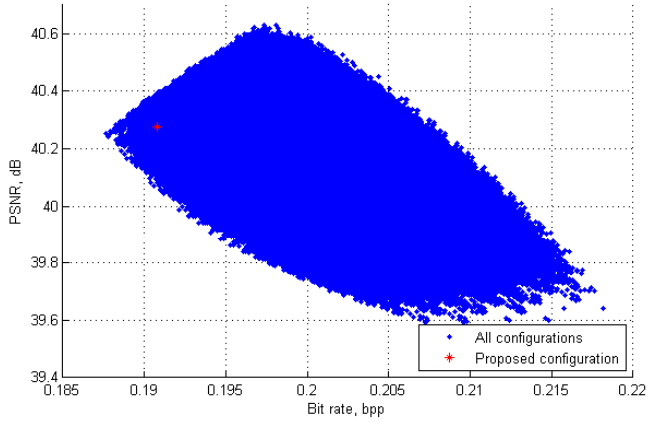



**Figure B.15:** Performance of different configurations for the sequence "News", 80% complexity

# Appendix C

# Examples of test video sequences

**Figure C.1:** Sequence “Akiyo”, 1st frame

**Figure C.2:** Sequence "City", 1st frame

**Figure C.3:** Sequence "Coastguard", 90th frame

**Figure C.4:** Sequence “Crew”, 70th frame


**Figure C.5:** Sequence “Foreman”, 125th frame


**Figure C.6:** Sequence "Football (1)", 100th frame

**Figure C.7:** Sequence "Football (2)", 10th frame

**Figure C.8:** Sequence “Ice”, 1st frame


**Figure C.9:** Sequence “Mobile”, 80th frame


**Figure C.10:** Sequence "News", 160th frame

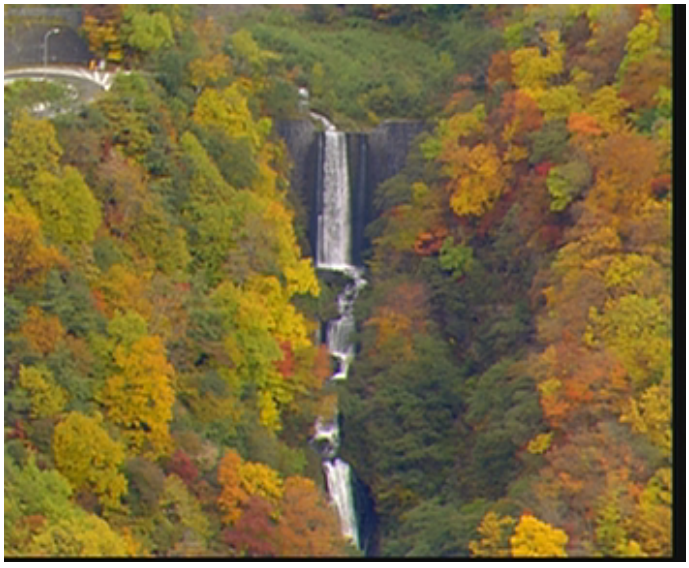


**Figure C.11:** Sequence "Waterfall", 200th frame

# List of Acronyms

| | |
|---|---|
| **3DVC** | 3D Video Coding |
| **ASO** | Arbitrary Slice Ordering |
| **AVC** | Advanced Video Coding |
| **CABAC** | Context-Adaptive Binary Arithmetic Coding |
| **CAVLC** | Context-Adaptive Variable-Length Coding |
| **CIF** | Common Intermediate Format |
| **CU** | Coding Unit |
| **DCT** | Discrete Cosine Transform |
| **DPCM** | Differential Pulse Code Modulation |
| **DVC** | Distributed Video Coding |
| **DWT** | Discrete Wavelet Transform |
| **EBCOT** | Embedded Block Coding with Optimal Truncation |
| **FMO** | Flexible Macroblock Ordering |
| **fps** | frame per second |
| **GB** | Gigabytes |
| **GOF** | Group Of Frames |
| **HD** | High-Definition |

| | |
|---|---|
| **HDTV** | High-Definition Television |
| **HEVC** | High Efficiency Video Coding |
| **ICT** | Information and Communications Technologies |
| **JPEG** | Joint Photographic Expert Group |
| **LAN** | Local Area Network |
| **LDPC** | Low-Density Parity-Check |
| **MC** | Motion Compensation |
| **ME** | Motion Estimation |
| **MJPEG** | Motion JPEG |
| **MOS** | Mean Opinion Score |
| **MSSIM** | Mean Structural Similarity Index |
| **MVC** | Multiview Video Coding |
| **PPSNR** | Perceptual PSNR |
| **P-R-D** | Power-Rate-Distortion |
| **PSNR** | Peak Signal-to-Noise Ratio |
| **R-D-C** | Rate-Distortion-Complexity |
| **RDO** | Rate-Distortion Optimization |
| **RGB** | Red, Green and Blue |
| **RLE** | Run Length Encoding |
| **RMSE** | Root Mean Square Error |
| **RoF** | Radio over Fiber |
| **ROI** | Region Of Interest |
| **RRC** | Radio Resource Control |

| | |
|---|---|
| **SLEP** | Systematic Lossy Error Protection |
| **SNR** | Signal-to-Noise Ratio |
| **SSIM** | Structural Similarity Index |
| **SVC** | Scalable Video Coding |
| **TDWZ** | Transform Domain Wyner-Ziv |
| **UEP** | Unequal Error Protection |
| **UHD** | Ultra High-Definition |
| **UVLC** | Universal Variable-Length Coding |
| **VLC** | Variable-Length Coding |
| **VQM** | Video Quality Metric |
| **WLAN** | Wireless Local Area Network |
| **WMAN** | Wireless Metropolitan Area Network |
| **WPAN** | Wireless Personal Area Network |
| **WWAN** | Wireless Wide Area Network |

# Bibliography


[1] A. Ukhanova, J. Korhonen, and S. Forchhammer, "Objective assessment of the impact of frame rate on video quality," *International Conference on Image Processing (ICIP)*, 2012.

[2] A. Ukhanova, E. Belyaev, L. Wang, and S. Forchhammer, "Power consumption analysis of constant bit rate video transmission over 3G networks," *Computer Communications Journal*, vol. 35, no. 14 (Special issue: Wireless Green Communications and Networking), pp. 1695–1706, 2012.

[3] A. Lebedev, T. Pham, M. Beltrán, X. Yu, A. Ukhanova, R. Llorente, I. Monroy, and S. Forchhammer, "Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band," *Optics Express Journal*, vol. 19, no. 26, pp. B895–B904, 2011.

[4] A. Lebedev, T. Pham, M. Beltrán, X. Yu, A. Ukhanova, L. Deng, N. Gonzalez, R. Llorente, I. Monroy, and S. Forchhammer, "Optimization of high-definition video coding and hybrid fiber-wireless transmission in the 60 GHz band," *European Conference on Optical Communication (ECOC)*, 2011.

[5] L. Wang, A. Ukhanova, and E. Belyaev, "Power consumption analysis of constant bit rate data transmission over 3G mobile wireless networks," *11th International Conference on Telecommunications for Intelligent Transport Systems (ITST)*, 2011.

[6] A.Ukhanova, A. Sergeev, and S. Forchhammer, "Extending JPEG-LS for low-complexity scalable video coding," *IS&T/SPIE Electronic Imaging*, 2011.

[7] A.Ukhanova and S. Forchhammer, "Low-complexity JPEG-based progressive video codec for wireless video transmission," *International Conference on Ultra Modern Telecommunications (ICUMT)*, 2010.

[8] X. Huang, A. Ukhanova, A. Veselov, S. Forchhammer, and M. Gilmutdinov, "Scalable-to-lossless transform domain distributed video coding," *Multimedia Signal Processing (MMSP)*, 2010.

[9] A. Ukhanova, E. Belyaev, and S. Forchhammer, "Encoder power consumption comparison of distributed video codec and H.264/AVC in low-complexity mode," *The 18th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2010.

[10] E. Belyaev, A.Turlikov, and A. Ukhanova, "Low-latency video transmission over high-speed WPANs based on low-power compression," *IEEE Wireless Communications & Networking Conference (WCNC)*, 2010.

[11] X. Huang, A. Ukhanova, E. Belyaev, and S. Forchhammer, "Temporal scalability comparison of the H.264/SVC and distributed video codec," *International Conference on Ultra Modern Telecommunications (ICUMT)*, 2009.

[12] "The world in 2011. ICT facts and figures." [Online]. Available: `http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf`

[13] "Global mobile data traffic forecast update, 2011-2016." [Online]. Available: `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf`

[14] Skype. [Online]. Available: `http://www.skype.com`

[15] YouTube. [Online]. Available: `http://www.youtube.com`

[16] Int. Telecommun. Union-Telecommun. (ITU-T) and Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC) JTC 1, ITU-T Rec.

H.264 & ISO/IEC 14496-10, "Advanced Video Coding for generic audiovisual services," 2003.

[17] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003.

[18] Int. Standards Org./Int. Electrotech. Comm., ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, "Introduction to SVC extension of Advanced Video Coding," 2005. [Online]. Available: `http://www.chiariglione.org/mpeg/technologies/mp04-svc/svc/`

[19] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, 2007.

[20] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, "Joint draft 8 of multiview video coding, Joint Video Team (JVT) doc. JVT-AB204," 2008.

[21] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626 – 642, 2011.

[22] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[23] B. Bross and W.-J. Han and J.-R. Ohm and G. J. Sullivan and T. Wiegand (Editors), "JCTVC-H1003: High Efficiency Video Coding (HEVC) text specification draft 7," 2012.

[24] H. Sohn, H. Yoo, W. D. Neve, C. S. Kim, and Y. M. Ro, "Full-reference video quality metric for fully scalable and mobile svc content," *IEEE Trans. on Broadcasting*, vol. 56, no. 3, pp. 269–280, 2010.

[25] R. Feghali, D. Wang, F. Speranza, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and

frame rate," *IEEE Trans. on Broadcasting*, vol. 53, no. 1, pp. 441–446, 2007.

[26] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 3, 2011.

[27] Y. Peng and E. Steinbach, "A novel full-reference video quality metric and its application to wireless video transmission," *International Conference on Image Processing (ICIP)*, 2011.

[28] "IEEE 802.15 WPAN millimeter wave alternative PHY task group 3c (TG3c)," 2009. [Online]. Available: `http://www.ieee802.org/15/ pub/TG3c.html`

[29] "WirelessHD." [Online]. Available: `http://www.wirelesshd.org`

[30] Standard ECMA-387, "High rate 60 GHz PHY, MAC and HDMI PALs."

[31] "DISCOVER codec." [Online]. Available: `www.discoverdvc.org`

[32] CVS repository for SVC software: JSVM. [Online]. Available: `:pserver:jvtuser:jvt.Amd.2@garcon.ient.rwth-aachen.de:/cvs/jvt`

[33] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *SPIE Visual Communications and Image Processing*, pp. 520–528, 2004.

[34] J. Korhonen, U. Reiter, and J. You, "Subjective comparison of temporal and quality scalability," *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2011.

[35] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*, 2nd ed. CRC Press, Inc., 2008.

[36] E. Belyaev, A. Turlikov, and A. Ukhanova, "Adaptive arithmetic coding in JPEG2000 standard," *Information-control systems*, vol. 31, no. 6, pp. 28–33, 2007, in Russian.

[37] D. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[38] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 620–636, 2003.

[39] Int. Telecommun. Union-Telecommun. (ITU-T), Recommendation H.261, "Video codec for audiovisual services at $p \times 64$ kbit/s," 1993.

[40] Int. Standards Org. / Int. Electrotech. Comm. (ISO/IEC) JTC 1, ISO/IEC 11 172-2 (MPEG-1), "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 2: Video," 1993.

[41] Int. Telecommun. Union-Telecommun. (ITU-T) and Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC) JTC 1, Recommendation H.262 and ISO/IEC 13 818-2 (MPEG-2 Video), "Generic coding of moving pictures and associated audio information – Part 2: Video," 1994.

[42] Int. Telecommun. Union-Telecommun. (ITU-T), Recommendation H.263, "Video coding for low bit rate communication," 2000.

[43] Int. Standards Org. and Int. Electrotech. Comm., (ISO/IEC) JTC 1, ISO/IEC 14 496-2 (MPEG-4 visual version 1), "Coding of audio-visual objects – part 2: Visual," 2003.

[44] Int. Telecommun. Union - Radiocommunication Sector, Recommendation ITU-R BT. 500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.

[45] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision.* MIT Press, 1993, pp. 207–220.

[46] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[47] Int. Telecommun. Union - Telecommunication Standardization Sector, ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," 1999.

[48] Y. Zhao, *Complexity Management for Video Encoders.* Ph.D. thesis, 2004, carried out at Robert Gordon University, UK.

[49] P. Lambert, W. D. Neve, P. D. Neve, I. Moerman, P. Demeester, and R. V. de Walle, "Rate-distortion performance of H.264/AVC compared to state-of-the-art video codecs," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 134–140, 2006.

[50] P. Amon, T. Rathgen, and D. Singer, "File format for scalable video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1174–1185, 2007.

[51] E. Belyaev, V. Grinko, and A. Ukhanova, "Power saving control for the mobile DVB-H receivers based on H.264/SVC standard," *The 8th Wireless Telecommunication Symposium*, 2009.

[52] V. Iverson, J. McVeigh, and B. Reese, "Real-time H.264/AVC codec on Intel architectures," *International Conference on Image Processing (ICIP)*, pp. 757–760, 2004.

[53] M. Chien, J. Huang, and P. Chang, "Complexity control for H.264 video encoding over power-scalable embedded systems," *IEEE International Symposium on Consumer Electronics(ISCE)*, pp. 221–224, 2009.

[54] R. de Queiroz, R. Ortis, A. Zaghetto, and T. Fonseca, "Fringe benefits of the H.264/AVC," *International Telecommunications Symposium*, pp. 166–170, 2006.

[55] Int. Telecommun. Union-Telecommun. (ITU-T) and Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC), ITU-T Recommendation T.81 and ISO/IEC International Standard 10918-1, "Information technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines," 1992.

[56] W. B. Pennebaker and J. L. Mitchell, *JPEG - Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

[57] L. Chen, N. Shashidhar, and Q. Liu, "Scalable secure MJPEG video streaming," *26th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 111–115, 2012.

[58] Int. Telecommun. Union-Telecommun. (ITU-T) and Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC) JTC 1, ITU-T Recommendation T.800 and ISO/IEC 15444-1, JPEG 2000 Part 1, "JPEG 2000 image coding system: Core coding system," 2000.

[59] D. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2001.

[60] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.

[61] Int. Telecommun. Union-Telecommun. (ITU-T), "T.802 : Information technology – JPEG 2000 image coding system: Motion JPEG 2000," 2005.

[62] ——, "T.832 : Information technology - JPEG XR image coding system - Image coding specification," 2009.

[63] Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC) 29199-2:2010, "Information technology - JPEG XR image coding system - Part 2: Image coding specification," 2010.

[64] I. E. Richardson, *The H.264 Advanced Video Compression Standard*. John Wiley & Sons, 2010.

[65] ISO/IEC-JTC1/SC29/WG11, Doc. JCTVC-E447, "HEVC reference software manual," 2011, ed. Geneva, Switzerland.

[66] J. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[67] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[68] A.Netravali and J.Limb, "Picture coding: a review," *Proceedings of the IEEE*, vol. 68, no. 3, pp. 366–406, 1980.

[69] G. J. Sullivan and T. Wiegand, "Video compression – from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, pp. 18–31, 2005.

[70] B. Li, G. J. Sullivan, and J. Xu, "Comparison of compression performance of HEVC working draft 4 with AVC High Profile," *Input Document to JCT-VC JCTVC-G399_r2*, 2011.

[71] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, 1965.

[72] ITU-T WP3/16 Media Coding and ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, "Joint preliminary call for proposals on Scalable Video Coding extensions of High Efficiency Video Coding (HEVC)," 2012.

[73] H. Choi, J. Nam, D. Sim, and I. Bajic, "Scalable video coding based on High Efficiency Video Coding (HEVC)," *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, pp. 346 – 351, 2011.

[74] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, 2011.

[75] B. Bartczak, P. Vandewalle, O. Grau, G. Briand, J. Fournier, P. Kerbiriou, M. Murdoch, M. Muller, R. Goris, R. Koch, and R. v. Vleuten, "Display-independent 3D-TV production and delivery using the layered depth video format," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 477–490, 2011.

[76] Y. Liu, F. Wu, and K. N. Ngan, "3D object-based scalable wavelet video coding with boundary effect suppression," *IEEE Interna-*

*tional Symposium on Circuits and Systems (ISCAS)*, pp. 1755–1758, 2007.

[77] C.-W. Deng and B.-J. Zhao, "Scalable 3D wavelet video coding scheme," *International Conference on Signal Processing (ICSP)*, pp. 68–71, 2008.

[78] H. Karim, C. Hewage, S. Worrall, and A. Kondoz, "Scalable multiple description video coding for stereoscopic 3D," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 2, pp. 745–752, 2008.

[79] L. Favalli and M. Folli, "A scalable multiple description scheme for 3D video coding based on the interlayer prediction structure," *International Journal of Digital Multimedia Broadcasting*, 2010.

[80] H.264/AVC JM reference software. [Online]. Available: `http://iphome.hhi.de/suehring/tml/`

[81] P. Dragotti and M. Gastpar, *Distributed Source Coding: Theory, Algorithms and Applications*. Academic Press, 2009.

[82] A. Reibman and B. Haskell, "Constraints on variable bit-rate video for ATM networks," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 2, no. 4, pp. 361–372, 1992.

[83] C.-Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios*, vol. 17, pp. 756–773, 1999.

[84] J.-S. Lee, F. D. Simone, T. Ebrahimi, N. Ramzan, and E. Izquierdo, "Quality assessment of multidimensional video scalability," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 38 – 46, 2012.

[85] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. on Multimedia*, vol. 10, no. 7, pp. 1316–1324, 2008.

[86] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE Trans. on Multimedia*, vol. 8, no. 5, pp. 973–980, 2006.

[87] Q. Huynh-Thu and M. Ghanbari, "Temporal aspect of perceived quality of mobile video broadcasting," *IEEE Trans. on Broadcasting*, vol. 54, no. 3, pp. 641–651, 2008.

[88] G. Yadavalli, M. Masry, and S. S. Hemami, "Frame rate preferences in low bit rate video," *International Conference on Image Processing (ICIP)*, 2003.

[89] J. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth: Comparing the effects of quantization vs. frame rate for streamed video," *Proc. of ACM CHI Hum. Factors Comput. Syst.*, pp. 535–542, 2004.

[90] N. van den Ende, H. de Hesselle, and L. Meesters, "Towards content-aware coding: User study," *European Conf. Interactive TV*, 2007.

[91] Y.-F. Ou, Z. Ma, and Y. Wang, "A novel quality metric for compressed video considering both frame rate and quantization artifacts," *International Workshop on Image Processing and Quality Metrics for Consumer (VPQM)*, 2009.

[92] Y.-F. Ou, Y. Zhou, and Y. Wang, "Perceptual quality of video with frame rate variation: A subjective study," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2446–2449, 2010.

[93] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202, "Joint scalable video model," 2007.

[94] N. C. Silver and W. P. Dunlap, "Should Fisher's z transformation be used?" *Journal of Applied Psychology*, vol. 72, no. 1, pp. 146–148, 1987.

[95] F. Dufaux, W. Gao, S. Tubaro, and A. Vetro, "Distributed video coding: Trends and perspectives," *EURASIP J. Image and Video Processing*, 2009.

[96] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: Selecting the most promising

application scenarios," *Signal Processing: Image Communication*, vol. 23, no. 5, pp. 339–352, 2008.

[97] N. Cheung, C. Tang, A. Ortega, and C. S. Raghavendra, "Efficient wavelet-based predictive Slepian-Wolf coding for hyperspectral imagery," *EURASIP Journal on Signal Processing - Special Issue on Distributed Source Coding*, pp. 3180–3195, 2006.

[98] A. Abrardo, M. Barni, E. Magli, and F. Nencini, "Error-resilient and low-complexity onboard lossless compression of hyperspectral images by means of distributed source coding," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1892–1904, 2010.

[99] F. Jelinek, "Buffer overflow in variable-length coding of fixed rate sources," *IEEE Trans. on Information Theory*, vol. 14, pp. 490–501, 1968.

[100] IEEE 802.3-2008, "IEEE Standard for information technology - Telecommunications and information exchange between systems - Local and Metropolitan area networks - Specific requirements - Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications," 2008.

[101] The Institute of Electrical and Electronics Engineers, Inc. IEEE Std 802.11, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," 1999.

[102] C. D. Knutson and J. M. Brown, *IrDA Principles And Protocols*. MCL Press, 2004.

[103] IEEE 802.15.1-2005, "IEEE Standard for information technology - Telecommunications and information exchange between systems - Local and Metropolitan area networks - Specific requirements - Part 15.1: Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPANs)," 2005.

[104] IEEE 802.15.4-2011, "IEEE Standard for local and metropolitan area networks - Part 15.4: Low-rate wireless personal area networks (LR-WPANs)," 2011.

[105] ZigBee Alliance, "Zigbee 2007," 2007.

[106] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan area networks. Part 16: Air interface fixed broad-band wireless access systems," 2004.

[107] ETSI Technical report 101 111 V3.0.1 (UMTS 21.01 version 3.0.1), "Requirements for the UMTS Terrestrial Radio Access system (UTRA)."

[108] Recommendation ITU-R M.1645, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000."

[109] Report ITU-R M.2134, "Requirements related to technical performance for IMT-advanced radio interface(s)."

[110] "Wireless Gigabit Alliance." [Online]. Available: `http://wirelessgigabitalliance.org`

[111] H. Singh, J. Oh, C. Kweon, X. Qin, H.-R. Shao, and C. Ngo, "A 60 GHz wireless network for enabling uncompressed video communication," *IEEE Communications Magazine*, 2008.

[112] Z. Jia, H.-C. Chien, Y.-T. Hsueh, A. Chowdhury, J. Yu, and G.-K. Chang, "Wireless HD services over optical access systems: Transmission, networking, and demonstration," *Conference on Optical Fiber Communication*, pp. 1–5, 2009.

[113] A. Belogolovy, E. Belyaev, A. Sergeev, and A. Turlikov, "Video compression for wireless transmission: reducing the power consumption of the WPAN hi-speed systems," *NEW2AN/ruSMART 2009*, vol. LNCS 5764, pp. 313–322, 2009.

[114] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 657–673, 2003.

[115] F. Fitzek, A. Kopsel, A. Wolisz, M. Krishnam, and M. Reisslein, "Providing application-level QoS in 3G/4G wireless systems: a comprehensive framework based on multi-rate CDMA," *IEEE Wireless Communications*, vol. 9, no. 2, pp. 42–47, 2002.

[116] C.-Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios*, vol. 17, pp. 756–773, 1999.

[117] W. Pu, Y. Lu, and F. Wu, "Joint power-distortion optimization on devices with MPEG-4 AVC/H.264 codec," *IEEE International Conference on Communications (ICC)*, 2006.

[118] N. Sklavos and K. Touliou, "A system-level analysis of power consumption and optimizations in 3G mobile devices," *Proceedings of the 1st International Conference on NewTechnologies, Mobility and Security (NTMS)*, 2007.

[119] J. Adams and G. Muntean, "Power save adaptation algorithm for multimedia streaming to mobile devices," *IEEE International Conference on Portable Information Devices*, 2007.

[120] G. P. Perrucci, *Energy Saving Strategies on Mobile Devices.* Ph.D. thesis, 2009, carried out at Aalborg University, Denmark.

[121] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, 2011.

[122] J. Korhonen and Y. Wang, "Power-efficient streaming for mobile terminals," *Proceedings of the International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 2005.

[123] E. de Diego Balaguer, F. Fitzek, O. Olsen, and M. Gade, "Performance evaluation of power saving strategies for DVB-H services using adaptive MPE-FEC decoding," *IEEE 16th International Symposium on Indoor and Mobile Radio Communications*, vol. 4, pp. 2221–2226, 2005.

[124] S. Loncaric, S. Grgic, , and B. Zovko-Cihlar, "Minimizing power consumption of the DVB-H receiver," *48th International Symposium ELMAR-2006 focused on Multimedia Signal Processing and Communications*, pp. 309–313, 2006.

[125] European Telecommunication Standard ETSI EN 302 304 v1.1.1 (2004-11), "Digital Video Broadcasting (DVB): Transmission system for handheld terminals (DVB-H)."

[126] E. Belyaev, T. Koski, J. Paavola, A. Turlikov, and A. Ukhanova, "Adaptive power saving on the receiver side in digital video broadcasting systems based on progressive video codecs," *The 11th International Symposium on Wireless Personal Multimedia Communications*, 2008.

[127] Z. He and S. K. Mitra, "From rate-distortion analysis to resource-distortion analysis," *IEEE Circuits and Systems Magazine*, vol. 5, no. 3, pp. 6–18, 2005.

[128] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 5, pp. 645–658, 2005.

[129] J. Kim, J. Kim, G. Kim, and C.-M. Kyung, "Power-rate-distortion modeling for energy minimization of portable video encoding devices," *IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1–4, 2011.

[130] X. Li, M. Wien, and J.-R. Ohm, "Rate-complexity-distortion optimization for hybrid video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 957–970, 2011.

[131] B. Foo, Y. Andreopoulos, and M. van der Schaar, "Analytical rate-distortion-complexity modeling of wavelet-based video coders," *IEEE Trans. on Signal Processing*, vol. 56, no. 2, pp. 797–815, 2008.

[132] R. Vanam, E. Riskin, S. Hemami, and R. Ladner, "Distortion-complexity optimization of the H.264/MPEG-4 AVC encoder using

the GBFOS algorithm," *Data Compression Conference (DCC)*, pp. 303–312, 2007.

[133] Z. He, W. Cheng, and X. Chen, "Energy minimization of portable video communication devices based on power-rate-distortion optimization," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 5, pp. 596–608, 2008.

[134] X. Lu, Y. Wang, and E. Erkip, "Power efficient h.263 video transmission over wireless channels," *International Conference on Image Processing (ICIP)*, 2002.

[135] J. Støttrup-Andersen, S. Forchhammer, and S. Aghito, "Rate-distortion-complexity optimization of fast motion estimation in H.264/MPEG-4 AVC," *International Conference on Image Processing (ICIP)*, 2004.

[136] G. Corrêa, P. Assuncao, L. Agostini, and L. da Silva Cruz, "Complexity control of high efficiency video encoders for power-constrained devices," *IEEE Trans. on Consumer Electronics*, vol. 57, no. 4, pp. 1866–1874, 2011.

[137] M. J. Osborne, *An Introduction to Game Theory*. Oxford University Press, 2003.

[138] S. Milani and G. Calvagno, "A game theory based classification for distributed downloading of multiple description coded videos," *International Conference on Image Processing (ICIP)*, pp. 3077–3080, 2009.

[139] ——, "Improving quality-of-experience for multiple description video transmission in peer-to-peer networks," *ACM Workshop on Advanced Video Streaming Techniques for Peer-to-Peer Networks and Social Networking*, pp. 65–69, 2009.

[140] J. Zhang, D. Wu, S. Ci, H. Wang, and A. Katsaggelos, "Power-aware mobile multimedia: a survey," *Journal of Communications*, vol. 4, no. 9, pp. 600–613, 2009.

[141] T. Wiegand, L. Noblet, and F. Rovati, "Scalable video coding for IPTV services," *IEEE Trans. on Broadcasting*, vol. 55, no. 2, pp. 527–538, 2009.

[142] N. Ramzan and E. Izquierdo, "Scalable and adaptable media coding techniques for future internet," in *The future internet.* Springer-Verlag, 2011, pp. 381–389.

[143] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalable video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1204–1217, 2007.